



Interactions de la région C-terminale de MLH1 nécessaires à la voie de réparation des mésappariements de l'ADN

Emeric Gueneau

► To cite this version:

Emeric Gueneau. Interactions de la région C-terminale de MLH1 nécessaires à la voie de réparation des mésappariements de l'ADN. Sciences agricoles. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA114804 . tel-00628928

HAL Id: tel-00628928

<https://theses.hal.science/tel-00628928>

Submitted on 4 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PARIS-SUD 11

FACULTÉ DE PHARMACIE DE CHÂTENAY-MALABRY

ECOLE DOCTORALE :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

ANNÉE 2010 - 2011

SÉRIE DOCTORAT N°

THÈSE

Présentée

À L'UNITÉ DE FORMATION ET DE RECHERCHE

FACULTE DE PHARMACIE DE CHATENAY-MALABRY

UNIVERSITÉ PARIS-SUD 11

pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-SUD 11

par

Emeric GUENEAU

Titre de la thèse :

INTERACTIONS DE LA REGION C-TERMINALE DE MLH1 NECESSAIRES A LA VOIE DE REPARATION DES MESAPPARIEMENTS DE L'ADN

Soutenue le 18 mars 2011

JURY :	Prof. Herman VAN TILBEURGH	Président
	Dr. Ivan MATIC	Rapporteur
	Dr. Marc DELARUE	Rapporteur
	Dr. Sophie QUEVILLON-CHERUEL	Examineur
	M. Etienne ROULEAU	Examineur
	Dr. Jean-Baptiste CHARBONNIER	Directeur de thèse

Avant-propos

Cette thèse, financée par le Ministère de l'Enseignement Supérieur et de la Recherche et la Fondation pour la Recherche Médicale, s'est déroulée au Laboratoire de Biologie Structurale et Radiobiologie au sein de l'iBiTec-S du CEA-Saclay. Je remercie Bernard GILQUIN, directeur du laboratoire, pour m'avoir accueilli au cours de ces trois années et demie.

La direction de ce travail a été assurée par Jean-Baptiste CHARBONNIER, qui m'a encadré et conseillé au quotidien. Sa disponibilité, son dynamisme et sa bienveillance m'ont permis de surmonter les difficultés rencontrées au cours de cette thèse. Par ailleurs, par sa rigueur et son esprit critique, il a su apporter à ce travail le recul nécessaire. Pour tout cela, je lui suis très reconnaissant.

Je tiens à remercier Herman VAN TILBEURGH pour avoir accepté de présider le jury de ma soutenance de thèse. J'exprime ma gratitude à Marc DELARUE et Ivan MATIC qui ont accepté de juger ce travail en tant que rapporteurs, ainsi qu'à Sophie QUEVILLON-CHERUEL et Etienne ROULEAU qui ont accepté de faire partie du jury. Je les remercie tous pour le temps qu'ils ont consacré à la lecture et à la critique de ce manuscrit.

Par ailleurs, cette thèse est loin d'être le résultat d'un travail isolé, je souhaiterais donc remercier toutes les personnes ayant contribué à son aboutissement.

Je tiens à exprimer ma profonde reconnaissance à Floriana LONDINO et Pierre BONNESOEUR qui m'ont épaulé tout au long de ma thèse. Sans eux, une grande partie des résultats n'aurait pas vu le jour. Je remercie également le programme "Plasticité et Instabilité du Génome" du CEA qui les a financés.

Je remercie également Marcela NUNEZ et Carine TELLIER-LEBEGUE pour leur participation au démarrage du projet.

Je tiens à exprimer toute ma gratitude à Marie-Hélène LE DU pour son aide en cristallographie ainsi que Sophie ZINN-JUSTIN pour sa contribution dans la modélisation.

Je suis également très reconnaissant envers les personnes avec lesquelles j'ai eu le plaisir de collaborer.

Je remercie Serge BOITEUX et Claudine DHERIN pour leur immense contribution au projet, aussi bien pour les expériences de génétiques que pour leurs nombreux conseils et idées.

Je remercie Pierre LEGRAND et Andy THOMPSON pour m'avoir accueilli plusieurs mois sur la ligne Proxima 1 et pour avoir passé de nombreuses heures à m'aider dans l'étude structurale.

Je remercie Simona MIRON pour son aide précieuse et ses conseils avisés en calorimétrie et en dichroïsme circulaire.

Enfin, je remercie Josan A. MARQUEZ et son équipe pour les nombreux essais de cristallisation réalisés sur leur plateforme et pour la qualité de leur collaboration.

Je ne saurai oublier dans ces quelques lignes toutes les personnes du laboratoire, permanents et doctorants, qui ont rendu ces trois années agréables. Mille mercis à tous.

Cela va de soi, je remercie ma famille pour son irremplaçable et inconditionnel soutien.

Enfin, si je n'avais qu'une personne à remercier, ce serait celle qui partage ma vie et qui a largement contribué à l'aboutissement de ce travail. Alors mille mercis, Caroline, pour m'avoir soutenu et supporté dans les moments difficiles, pour ta patience et tes nombreuses attentions.

Sommaire

NOMENCLATURE	8
INTRODUCTION GENERALE	11
1. REVUE DE LITTERATURE	14
1.1 LES PROTEINES MUTL DANS LA REPARATION DES MESAPPARIEMENTS	14
1.1.1 LA VOIE DE REPARATION DES MESAPPARIEMENTS	14
1.1.2 MUTL ET SES HOMOLOGUES	46
1.2 CONTEXTE DU PROJET: ETUDE STRUCTURALE ET FONCTIONNELLE DE LA REGION C-TERMINALE DES HOMOLOGUES EUCARYOTES DE MUTL	55
1.2.1 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE MUTL	55
1.2.2 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE PMS2: ACTIVITE ENDONUCLEASE DU DOMAINE C-TERMINAL	57
1.2.3 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE MLH1	61
1.3 OBJECTIFS DU PROJET: CARACTERISATION BIOPHYSIQUE ET STRUCTURALE DE LA REGION C-TERMINALE DE MUTL α	63
1.3.1 CARACTERISER LES SITES D'INTERACTIONS DU DOMAINE C-TERMINAL DE MUTL α	63
1.3.2 RESOLUTION DE LA 1 ^{ERE} STRUCTURE EUCARYOTE DU DOMAINE C-TERMINAL DE MUTL α	64
2. MATERIEL ET METHODES	67
2.1 EXPRESSION ET PURIFICATION DES PROTEINES	67
2.1.1 CONSTRUCTION DES DOMAINES	67
2.1.2 EXPRESSION ET PURIFICATION DES REGIONS C-TERMINALES DE SCMLH1, HSMLH1 ET SCMLH1*SCPMS1	67
2.2 CARACTERISATION BIOPHYSIQUE DES PROTEINES PURIFIEES	72
2.2.1 CARACTERISATION DU REPLIEMENT PAR DICHROÏSME CIRCULAIRE	72
2.2.2 CARACTERISATION DE LA POLYDISPERSITE PAR DIFFUSION DYNAMIQUE DE LA LUMIERE	75
2.2.3 CARACTERISATION DES INTERACTIONS PAR CALORIMETRIE A TITRAGE ISOTHERME	76
2.3 ETUDE DES STRUCTURES TRIDIMENSIONNELLES	79
2.3.1 MODELISATION PAR HOMOLOGIE DU DOMAINE C-TERMINAL DU COMPLEXE MLH1*Pms1	79
2.3.2 MODELISATION DU PEPTIDE MIP DE EXO1 SUR LA SURFACE DE LA REGION C-TERMINALE DE MLH1	80
2.3.3 LA CRISTALLOGRAPHIE AUX RAYONS X DES PROTEINES	82
3. ANALYSE FONCTIONNELLE ET STRUCTURALE DU DOMAINE C-TERMINAL DE MUTLα	97
3.1 IDENTIFICATION ET CARACTERISATION D'UN NOUVEAU SITE D'INTERACTION DE MLH1	97

3.1.1	IDENTIFICATION DU SITE D'INTERACTION DES PROTEINES CONTENANT UN MOTIF MIP PAR MUTAGENÈSE DIRIGÉE ET MODELISATION PAR HOMOLOGIE	97
3.1.2	PURIFICATION ET CARACTÉRISATION BIOPHYSIQUE DE LA RÉGION C-TERMINALE DE MLH1	103
3.1.3	CARACTÉRISATION PAR CALORIMÉTRIE DES INTERACTIONS DE PEPTIDES CONTENANT UN MOTIF MIP AVEC LA RÉGION C-TERMINALE DE MLH1	107
3.2	ANALYSE STRUCTURALE DU DOMAINE C-TERMINAL DE MutLα	118
3.2.1	ÉTUDE STRUCTURALE DU DOMAINE C-TERMINAL DE MLH1	118
3.2.2	RÉSOLUTION DE LA STRUCTURE CRISTALLOGRAPHIQUE DE LA RÉGION C-TERMINALE DU COMPLEXE MLH1*PMS1 DE <i>S. CEREVISIAE</i>	124
3.2.3	ANALYSE DE LA STRUCTURE CRISTALLOGRAPHIQUE DU DOMAINE C-TERMINAL DE MutL α DE <i>S. CEREVISIAE</i>	152
4.	CONCLUSIONS ET PERSPECTIVES	177
	ANNEXE	207
	ANNEXE 1: PRÉPARATION DE BACTÉRIES CHIMIOCOMPÉTENTES (INOUE ET AL., 1990)	207
	ANNEXE 2: COMPOSITION DU MILIEU DE CULTURE POUR LA PRODUCTION DE PROTÉINE MARQUÉE AU SÉLENIUM	210
	ANNEXE 3: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DES DOMAINES 485-769 ET 503-769 DE scMLH1	211
	ANNEXE 4: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DU DOMAINE 486-756 DE hscMLH1	215
	ANNEXE 5: SPECTRE DE DICHROISME CIRCULAIRE ET DÉNATURATION THERMIQUE DU VARIANT E682A DU DOMAINE C-TERMINAL DE MLH1	219
	ANNEXE 6: THERMOGRAMME ET ISOTHERME DU DOMAINE C-TERMINAL DE MLH1 ET DE PEPTIDES DÉRIVÉS DU MOTIF MIP	220
	ANNEXE 7: CARACTÉRISATION PAR BSI DES INTERACTIONS DE PEPTIDES CONTENANT UN MOTIF MIP AVEC LA RÉGION C-TERMINALE DE MLH1	224
	ANNEXE 8: CONDITIONS DE CRISTALLOGÈNESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1 (485-769)	225
	ANNEXE 9: CONDITIONS DE CRISTALLOGÈNESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1 (503-769)	226
	ANNEXE 10: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DES DOMAINES 485-769 DE scMLH1*635-873 DE scPMS1	228
	ANNEXE 11: CONDITIONS DE CRISTALLOGÈNESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1*PMS1	232
	ANNEXE 12: SCRIPTS DE REMPLACEMENT MOLÉCULAIRE	233

ANNEXE 13: COMPARAISON DES STUCTURES SECONDAIERS PREDITES ET REELLES DES CTD DE MLH1 ET PMS1	237
ANNEXE 14: ANALYSE DE L'INTERFACE DE DIMERISATION PAR LE SERVEUR PROTORP ; RESIDUS DE MLH1 A LA SURFACE DE DIMERISATION	238
ANNEXE 15: ANALYSE DE L'INTERFACE DE DIMERISATION PAR LE SERVEUR PROTORP ; RESIDUS DE PMS1 A LA SURFACE DE DIMERISATION	239
ANNEXE 16: ALIGNEMENTS DE 10 SEQUENCES DU DOMAINE C-TERMINAL DE MLH1 A L'AIDE DU PROGRAMME KALIGN	240
ANNEXE 17: ALIGNEMENTS DE 10 SEQUENCES DU DOMAINE C-TERMINAL DE PMS1 ET 2 SEQUENCES DU DOMAINE C-TERMINAL DE MLH3 A L'AIDE DU PROGRAMME KALIGN	242

Nomenclature

ADP	Adénosine Di-Phosphate
ADP_nP ou AMPPNP	Adenosine 5'-[β,γ -imido]triphosphate
AID	Activation Induced cytidine Deaminase
ATP	Adénosine Tri-Phosphate
ATPγS	Adenosine 5'-O-thiotriphosphate
ATR	Ataxia Telangiectasia and Rad3-related protein
BER	Base Excision Repair
CD	Circular Dichroism
CHARMM	Chemistry at HARvard Molecular Mechanics
Co-IP	Co-ImmunoPrécipitation
CSR	Class Switch Recombination
CTD	C Terminal Domain
DLS	Dynamic Light Scattering
DMSO	DiMéthylSulfOxyde
DO600nm	Densité Optique à 600nm
EDTA	Acide EthylèneDiamine Tétra acétique
ERISCAM	Estimation des RISques de CAncer chez les porteurs de mutation des gènes MMR
FANCIJ	Fanconi anemia complementation group J
GST	Glutathione S-Transferase
HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazineéthanesulfonique
HNPCC	Hereditary NonPolyposis Colorectal Cancer
HTX	High Throughput Crystallisation
IDCL	Interdomain connecting loop
IDL	Insertion et délétion
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
InSiGHT	International Society for Gastrointestinal Hereditary Tumours
LB	Lysogeny Broth
LM-PCR	Ligation-Mediated Polymerase Chain Reaction
MBP	Maltose-Binding Protein
MIP	Mlh1 Interacting Protein
MLH	MutL Homolog protein
MMR	MisMatch Repair
MNNG	N-methyl-N'-nitro-N-nitrosoguanidine
MSH	MutS homolog protein
MSI	Microsatellite Instability

MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NER	Nucleotide Excision Repair
NHEJ	NonHomologous End Joining
NTD	N terminal domain
PBS	Phosphate Buffered Saline
PCNA	Proliferating Cell Nuclear Antigen
PEG	PolyEthylène Glycol
PIP	PCNA Interacting Protein box
PMS	Post Meiotic Segregation increased protein
PMSF	Fluorure de PhénylMéthaneSulfonyl
Ppi	Eau pour préparations injectables
PSI-BLAST	Position Spécific itéated BLAST
PSI-PRED	Position Spécific itéated PREDiction
PWWP	Motif Pro-Trp-Trp-Pro
RFC	Replication Factor C
RPA	Replication Protein A
SAXS	Small-Angle X-ray Scattering
SHM	Somatic Hypermutation
SSB	Single-strand binding protein
TEV	Tobacco Etch Virus protease
TFA	Acide TriFluoroAcétique
Tris	Trishydroxyméthylaminométhane
UNG	Uracil N-Glycosylase
UV	Unclassified variants
XDS	X-ray Detector Software
3R	Réplication, Recombinaison et Réparation

INTRODUCTION GENERALE

« Je viens de réaliser tardivement que l'ADN est tellement précieux qu'il doit probablement exister des mécanismes de réparation » (Crick, 1974). Par ces mots écrits en 1974 dans une rétrospective sur la découverte que lui et Watson avaient faite vingt et un ans plus tôt, Francis Crick met en avant la structure dynamique de la molécule d'ADN. Auparavant, l'ADN, qui renferme notre patrimoine génétique, était considéré comme une entité extrêmement stable. En fait, l'ADN possède un métabolisme regroupé sous la dénomination «3R» pour Réplication, Recombinaison et Réparation.

Ce métabolisme permet de transmettre la même information génétique d'une cellule à une autre et d'un organisme à un autre. Pour être transmis de façon fidèle, l'ADN doit comporter peu d'erreurs lors de la réplication et peu de dommages liés à l'action d'agents chimiques ou de rayonnements ionisants. On estime que l'ADN d'une cellule humaine subit en moyenne plusieurs dizaines de milliers de lésions par cellule et par jour, dans des conditions habituelles d'activité métabolique et d'exposition aux facteurs environnementaux. Bien que la mutagenèse soit un moteur de l'évolution, ces lésions sont bien souvent néfastes (cancer, vieillissement, etc.). Pour faire face à ces dommages, nos cellules possèdent de nombreuses stratégies connexes au sein des réseaux d'interactions des 3R.

Au sein de notre laboratoire, nous étudions différents réseaux d'interactions des 3R par des approches bio-informatiques, biochimiques et structurales. Un des projets du laboratoire porte sur le réseau effectué par la protéine eucaryote Mlh1 de la voie de réparation des mésappariements (MMR). Ce mécanisme de réparation consiste à remplacer une base qui n'est pas appariée correctement. L'existence de ce mécanisme de réparation des erreurs d'appariement a été prouvée au début des années 70 chez les bactéries et en 1985 chez la levure. En 1989, un système minimal de réparation des mésappariements, constitué de protéines bactériennes purifiées, a été reconstitué *in vitro*, démontrant que les gènes essentiels de ce mécanisme avaient été identifiés. Ce correcteur d'orthographe génétique fait intervenir en particulier la protéine Mlh1 et son homologue bactérien MutL. Le nom Mut est la contraction de «Mutateur» car les cellules déficientes en MMR accumulent des mutations et déséquilibrent la balance entre les suppresseurs de tumeurs et les oncogènes. Chez l'homme, une mutation au niveau d'un gène du MMR implique une prédisposition héréditaire au cancer du côlon, expliquant ainsi les cas récurrents de cancer au sein d'une même famille, observés dès 1913 par AS. Warthin.

Ce mécanisme reste encore énigmatique sous bien des aspects. La coordination des différentes étapes de réparation ainsi que la communication entre les différents

systèmes de réparation se chevauchant sont à ce jour peu compris. La coordination de nombreuses étapes au sein du MMR, semble impliquer la protéine MutL ou ses homologues eucaryotes. C'est pour ces différentes raisons que nous nous sommes intéressés au mécanisme moléculaire de Mlh1 mettant en œuvre de nombreuses interactions protéine-protéine, en utilisant une approche structurale et biochimique complétée par des approches génétiques de nos collaborateurs du laboratoire de S. Boiteux (CEA-CNRS, Fontenay-aux-Roses).

La première partie présente une revue bibliographique qui précise le contexte général de la réparation des mésappariements ainsi que les connaissances structurales et fonctionnelles concernant la protéine MutL et ses homologues eucaryotes. Cette partie se termine par les questions adressées au cours de cette thèse, autour du rôle des homologues de MutL chez les eucaryotes, et tout particulièrement de la fonction de leur domaine C-terminal.

La deuxième partie présente le matériel utilisé ainsi que les diverses méthodes expérimentales mises en œuvre au cours de ce projet. Nous détaillerons notamment, le puissant outil que sont les méthodes biophysiques pour comprendre les caractéristiques fonctionnelles.

La troisième partie rassemble l'ensemble des résultats de l'étude: dans un premier temps, la combinaison de méthodes biophysiques et génétiques a permis d'identifier un site d'interaction important pour les événements intervenant en aval de la reconnaissance du mésappariement; dans un second temps, nous montrerons comment la cristallographie aux rayons X a permis de résoudre la structure d'un domaine essentiel au bon fonctionnement du mécanisme de réparation MMR. Ces résultats sont discutés au fur et à mesure de leur présentation dans des sections « discussion » séparées

Enfin, la dernière partie reprendra les conclusions principales de cette thèse et proposera un certain nombre de perspectives correspondant à des travaux initiés et à des études qu'il nous semble important de mener dans le futur dans un laboratoire comme le LBSR.

1. REVUE DE LITTÉRATURE

1.1 LES PROTEINES MUTL DANS LA REPARATION DES MESAPPARIEMENTS

1.1.1 LA VOIE DE REPARATION DES MESAPPARIEMENTS

i. Vue globale de la réparation des mésappariements

Le mécanisme de réparation des mésappariements de l'ADN (MMR) joue un rôle important dans la maintenance de la stabilité du génome et permet ainsi de limiter la propagation de mutations. Le mécanisme MMR répare les erreurs de mésappariements introduites sur l'ADN. Il s'agit notamment des erreurs qui n'ont pas été corrigées par la fonction de relecture des ADN polymérases (ou "proofreading" en anglais). Certaines modifications des bases de l'ADN peuvent également être considérées comme des mésappariements; c'est le cas de l'appariement d'une base avec un uracile, généralement trouvé sous forme G·U. C'est aussi le cas de certains types de dommage comme l'oxydation d'une guanine en 8-oxo-G, cette dernière base s'appariant préférentiellement avec une adénosine; c'est enfin le cas de la méthylation d'une guanine suite à l'action d'agent alkylant, favorisant l'appariement avec une thymine plutôt qu'avec une cytosine (Kunz et al., 2009).

La spécificité du mécanisme MMR repose sur sa capacité à discriminer le brin néosynthétisé du brin parental. Ce système intervient préférentiellement sur les mésappariements post-réplicatifs où l'erreur est située sur le brin néosynthétisé. Le MMR est moins impliqué dans la réparation des mésappariements liés à une modification chimique de la base puisque ces dommages peuvent concerner l'un ou l'autre des brins d'ADN.

Ainsi, contrairement aux autres voies de réparation de l'ADN (Hoeijmakers, 2001), le mécanisme MMR doit prendre en charge une base chimiquement intacte mais ne formant pas une paire de type Watson-Crick. Ce mésappariement va alors modifier le réseau de liaisons hydrogène entre les bases azotées et provoque une différence d'enthalpie libre de l'ordre de 2-3kcal/mol par rapport à un appariement de type Watson-Crick (Friedberg, 2003). Cette différence d'énergie se traduirait à elle seule par une fréquence d'erreur de 1 à 10% par nucléotide, c'est-à-dire l'incorporation d'une base incorrecte tous les 10 à 100 nucléotides du même type (Kunkel, 2004).

Les mésappariements affectent la structure hélicoïdale de l'ADN mais de différentes manières selon le type de mésappariement. Les différents mésappariements possibles ne sont pas pris en charge avec la même efficacité: les mésappariements G·T et A·C sont corrigés de façon très efficace chez les procaryotes comme chez les eucaryotes; en revanche, le mésappariement C·C n'est pas réparé par le mécanisme MMR (Brown et al., 2001; Kramer et al., 1984). Il est intéressant de noter que ce

sont donc les mésappariements les plus stables (comme G·T et A·C) qui sont les mieux réparés alors que ceux qui adoptent une configuration énergétiquement défavorable ne sont pas reconnus. Si l'on considère le système MMR comme une sauvegarde de la relecture des polymérases, il est finalement intéressant d'observer que le MMR répare de manière plus efficace les mésappariements les mieux tolérés par les polymérases, *i.e.* ceux qui provoquent les plus faibles distorsions. L'action de cette voie de réparation permet alors d'abaisser le taux de mutations d'un facteur 100 à 1000 (Modrich and Lahue, 1996).

Le mécanisme MMR prend également en charge un second type de dommage à savoir les boucles d'insertion et de délétion (IDLs) (Parker and Marinus, 1992). Ces boucles apparaissent principalement au niveau de séquences appelées microsatellites. Les microsatellites sont composés de motifs mono- ou di-nucléotidiques répétés un certain nombre de fois. Lorsque les polymérases rencontrent ces séquences microsatellites, elles ont tendance à omettre de synthétiser un ou plusieurs nucléotides; on parle alors de glissement (ou "slippage" en anglais). Il en résulte l'apparition de boucles d'insertion/délétion (Loeb, 1994; Peltomaki, 2001). Un grand nombre de ces microsatellites sont répartis dans des régions non-codantes du génome mais certains existent aussi au sein de régions codantes de gènes humains (Jacob and Praz, 2002). L'introduction d'insertions-délétions dans ces gènes s'accompagne le plus souvent d'un changement de phase de lecture et de l'expression d'une protéine tronquée. Par analyse de lignées cellulaires associées à des cancers colorectaux, des altérations ont été découvertes pour des gènes suppresseurs de tumeurs comme le récepteur du TGF- β , *TGF β -RII* (Markowitz et al., 1995) qui possède une séquence poly-A₁₀ ou le gène Phosphate and TENsin homolog, *PTEN* qui contient une séquence poly-A₆ (Guanti et al., 2000). Des instabilités de microsatellites ont été observées pour des gènes codant des facteurs pro-apoptotiques tels que le gène *BAX* (pour Bcl-2 Associated X protein) et sa séquence poly-G₈ (Rampino et al., 1997) ou le gène *Caspase-5* et sa séquence poly-A₁₀ (Schwartz et al., 1999). Enfin des instabilités de microsatellites ont été rapportées dans des gènes de réparation de l'ADN tels que les gènes du MMR *hMSH3* {poly-A₈}, *hMSH6* {poly-C₈} (Malkhosyan et al., 1996) et *hMLH3* {poly-A₆; poly-A₈; poly-A₉} (Loukola et al., 2000), le gène *RAD-50* impliqué dans la réparation des cassures double-brin {poly-A₈; poly-A₉} (Ikenoue et al., 2001) ou le gène *BLM*, de la famille des hélicases RecQ et associé au syndrome de Bloom {poly-A₉} (Calin et al., 2001).

ii. Le mécanisme MMR chez *E. coli*

Le système MMR a été découvert chez *E. coli* par Glickman et Radman (Glickman and Radman, 1980) qui ont montré que l'inactivation des gènes *mut* (*mutS*, *mutL* et *mutH*) entraîne une augmentation du nombre de mutations (phénotype hypermutateur). Quelques années plus tard, l'équipe de P. Modrich a

montré l'implication des protéines MutS, MutL et MutH dans la réparation des mésappariements (Lu et al., 1983).

La réparation des mésappariements ou des boucles d'insertion/délétion (ou IDLs) se fait en trois étapes (**Figure 1 - 1**):

- 1) Reconnaissance d'un mésappariement ou d'un IDL par les protéines du système MMR;
- 2) Discrimination du brin néosynthétisé et du brin parental;
- 3) Excision du brin néosynthétisé et synthèse d'un nouveau brin exempt d'erreur la plupart du temps.

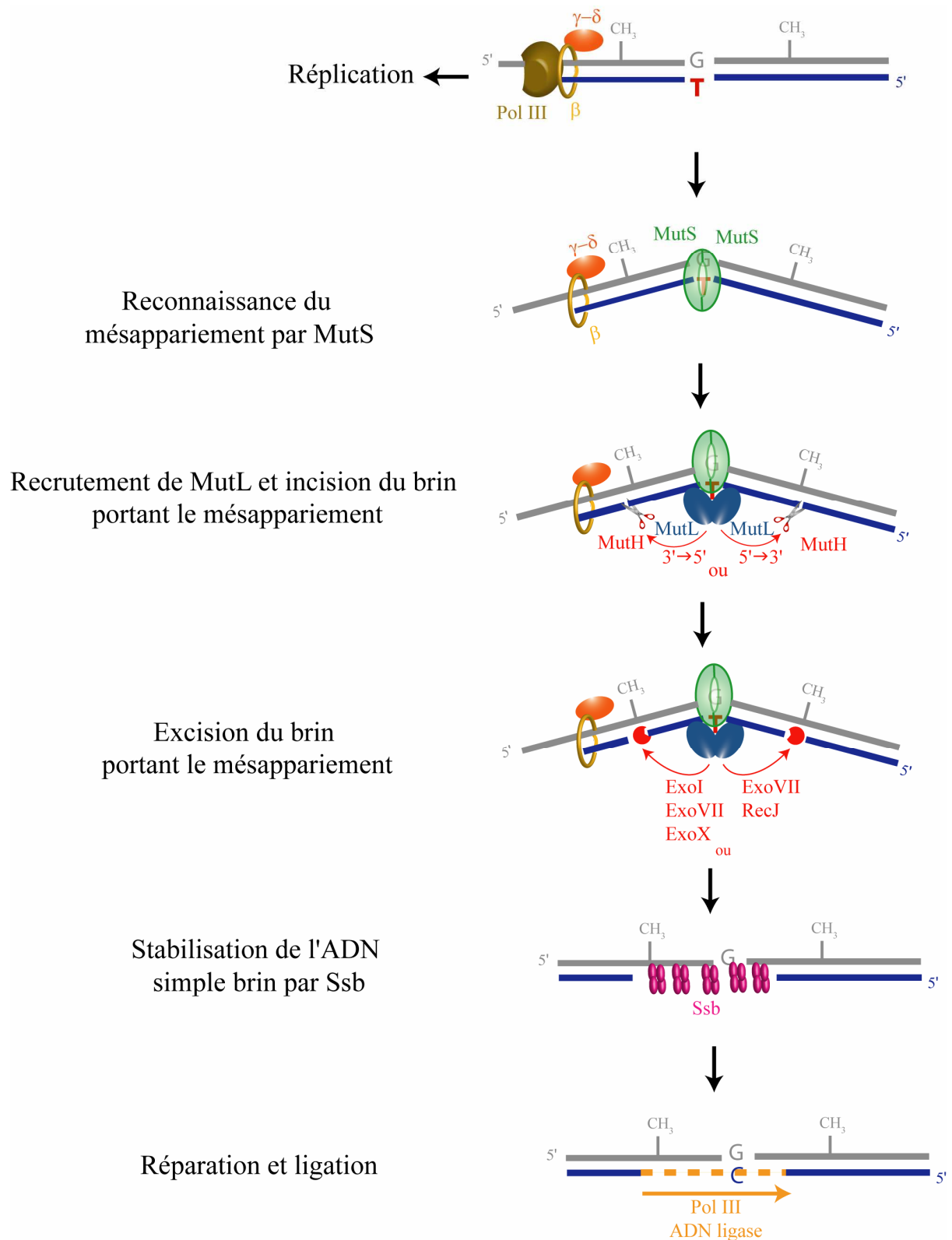


Figure 1 - 1 : Schéma de la réparation d'un mésappariement G-T par le MMR chez *E. coli* (d'après (Jiricny, 2006))

Reconnaissance du mésappariement sur le brin d'ADN

La protéine MutS reconnaît les mésappariements (hormis ceux de type C·C) et les IDLs. Les structures cristallographiques de MutS de *E. coli*, ecMutS, (Lamers et

al., 2000; Natrajan et al., 2003) et de *T. aquaticus*, taqMutS, (Obmolova et al., 2000) ont été déterminées en présence d'un mésappariement G·T et d'ADP pour la première et d'une insertion d'un nucléotide pour la seconde. Dans les deux cas, la protéine est sous forme homodimérique (**Figure 1 - 2(A)**): chaque monomère est composé de trois domaines fonctionnels, un domaine de reconnaissance du mésappariement, un domaine «pince» et un domaine ATPase.

Lors de la reconnaissance d'un mésappariement, les domaines de reconnaissance et «pince» de chaque monomère participent différemment. La structure de MutS lié à son substrat est considérée pour cette raison, comme un homodimère asymétrique où chaque chaîne joue un rôle différent (**Figure 1 - 2(B)**). Dans la structure d'ecMutS en complexe avec un ADN contenant un mésappariement G·T, le domaine «pince» du premier monomère interagit avec le squelette phosphodiester de l'ADN tandis que le domaine de reconnaissance du mésappariement du second monomère se lie spécifiquement au mésappariement *via* un motif conservé FxE. L'interaction de ce motif est de type π -stacking entre la phénylalanine F36 et la base de la thymine mésappariée. Les domaines «pince» et de reconnaissance, respectivement du premier et second monomère, sont, quant à eux, impliqués dans la stabilisation de la conformation globale de l'ADN.

Il n'y a pas à ce jour de structure de MutS lié à un homoduplex d'ADN et il est difficile de comprendre comment l'homodimère de MutS se lie spécifiquement au mésappariement. A cela s'ajoute le rôle que joue la fixation de nucléotide dans le site ATPase. Chaque dimère de MutS possède un site ATPase, chacun pouvant fixer une molécule d'ATP, d'ADP ou rester vide. Il a été proposé que ce site intervienne dans les changements conformationnels liés à la reconnaissance de l'erreur par un mécanisme encore mal compris. Cependant, des études dynamiques de microscopie à force atomique (Wang et al., 2003), d'échange de deutérium suivi par spectrométrie de masse (Mendillo et al., 2010) et d'analyses des modes normaux à basses fréquences (Mukherjee et al., 2009) couplées aux structures cristallographiques et aux études génétiques permettent d'envisager le scénario suivant:

- Liaison à l'ADN à la recherche d'un mésappariement: la structure cristallographique de MutS en l'absence d'ADN (Obmolova et al., 2000) montre que l'homodimère est dans une configuration ouverte où le domaine de fixation à l'ADN semble désordonné car invisible dans les cartes de densité électronique. Seul le domaine ATPase se structure pour former le dimère. Plusieurs études cinétiques semblent indiquer que MutS recherche un mésappariement *via* un mécanisme de collision aléatoire où le domaine de fixation à l'ADN se structure sur la double hélice. MutS lié à l'ADP formerait un complexe instable avec l'ADN correctement apparié, pouvant s'associer et se dissocier rapidement, le motif FxE sondant l'appariement de l'ADN dans ce court laps de temps (Jacobs-Palmer and Hingorani, 2007);

- Reconnaissance du mésappariement: A l'inverse, lorsque le dimère MutS*ADP rencontre un mésappariement, il se lie à ce dernier avec une affinité beaucoup plus forte. La transition pour passer d'une étape de recherche à une étape de reconnaissance implique un changement significatif depuis une configuration relativement lâche du domaine de fixation de l'ADN avec des interactions MutS*ADN de faible affinité, à une configuration où le dimère forme une interaction spécifique en courbant l'ADN d'environ 45-60° au niveau du mésappariement, conformément aux structures cristallographiques. Cette stabilisation s'effectue par l'interaction spécifique de la phénylalanine du motif FxE. Il a été proposé que la base mésappariée ou insérée puisse être extrudée de la double hélice pour stabiliser la torsion de l'ADN au niveau du mésappariement (Wang et al., 2003);
- Traduction du signal: des études cinétiques ont montré que chez ecMutL, cette reconnaissance spécifique induit un échange rapide d'une molécule d'ADP avec une molécule d'ATP (Acharya et al., 2003; Mendillo et al., 2010). Complexifiant le mécanisme, plusieurs études indiquent que les sites ATPases fonctionnent selon un mode asymétrique: la structure d'ecMutS lié au mésappariement G·T ne comporte qu'une molécule d'ADP (Lamers et al., 2000). Des études cinétiques ont montré que les deux sites ATPases n'ont pas la même affinité vis-à-vis de l'ATP (Antony and Hingorani, 2004). Il est alors possible que seul un des deux sites soit touché par l'échange de nucléotides, le second site restant fixé à une molécule d'ADP. Dans tous les cas, cet échange va induire une nouvelle configuration, non consommatrice d'énergie puisque l'ATP n'est pas hydrolysée. Des expériences d'empreinte sur l'ADN (ou "footprinting" en anglais) montrent que l'échange ADP→ATP réduit l'affinité entre MutS et le mésappariement (Joshi et al., 2000); l'homodimère de MutS va alors se transformer en pince glissante qui peut diffuser sur le duplex d'ADN et sert de signal d'assemblage du complexe de réparation.

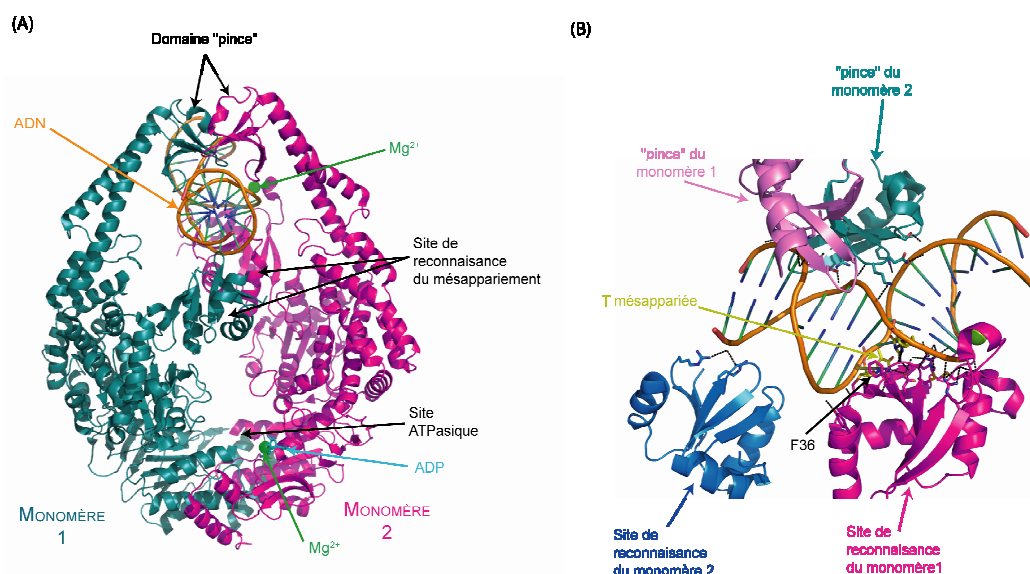


Figure 1 - 2: Structure du complexe MutS (A) Structure cristallographique (code PDB:1E3M) à 2.2Å d'ecMutS lié à une double hélice d'ADN portant un mésappariement G·T(en orange); les monomères 1 et 2 sont respectivement colorés en vert marin et magenta; les cations Mg^{2+} sont représentés par des sphères vertes (B) Détail du site de fixation d'ecMutS à une thymine mésappariée; le site de reconnaissance se liant spécifiquement à la thymine est représenté en magenta, le second site est représenté en bleu; les deux domaines pinces sont représentés en rose et vert marin

Coordination du MMR depuis la reconnaissance du mésappariement jusqu'à l'incision de l'ADN néosynthétisé au site hémiméthylé pouvant se trouver à plus de 1000 paires de base

Le complexe MutS*ADN activé par l'ATP interagit avec l'homodimère MutL (Selmane et al., 2003). Il a été observé par retard sur gel que MutL n'affecte ou n'interagit pas avec MutS lié à un mésappariement. Ceci suggère que MutS et MutL interagissent seulement qu'après l'échange $ADP \rightarrow ATP$ (Acharya et al., 2003).

Le triplex MutS*MutL*ADN active l'endonucléase MutH afin d'inciser le brin d'ADN portant le mésappariement. Chez *E. coli*, l'ADN est méthylé au niveau de l'azote N₆ des adénines. Mais au cours de la réplication de l'ADN, le brin néosynthétisé est transitoirement non-méthylé et c'est cette différence qui permet de discriminer les deux brins. Ainsi, l'endonucléase MutH est recrutée par le complexe MutS*MutL et incise le brin non-méthylé permettant l'excision du mésappariement. Ce site hémiméthylé peut se trouver à plus de 1000 paires de bases (ou bp pour base pair en anglais), du côté 3' ou 5' du mésappariement, faisant du MMR un mécanisme bidirectionnel (Iyer et al., 2006).

Une question, non élucidée encore à ce jour chez *E. coli*, est celle du mécanisme moléculaire qui permet d'associer la reconnaissance du mésappariement par les protéines MutS à l'incision du brin néosynthétisé par l'endonucléase MutH. Différents facteurs semblent importants dans cette coordination: les changements de

conformations ATP/ADP dépendant de MutS, le rôle d'intermédiaire de la protéine MutL par ses interactions protéine-protéine avec MutS et MutH et ses changements de conformations ATP/ADP dépendants. Trois modèles ont été proposés pour expliquer la communication entre le mésappariement et l'incision au niveau du site d'hémiméthylation (Kunkel and Erie, 2005):

- Le modèle de translocation (**Figure 1 - 3 à gauche**)

Le modèle de P. Modrich (Allen et al., 1997) est né de l'observation en microscopie électronique de boucles d'ADN dites α en présence de MutS, MutL, d'ATP et d'un mésappariement. Il a également été observé que la taille de ces boucles augmentant en fonction du temps. Dans 80% des cas, le mésappariement se trouve enfermé au sein de la boucle, à la base de laquelle se trouvent MutS et MutL. *A contrario*, ces boucles α ne se forment pas en l'absence de mésappariements, ni en présence d'un mésappariement C·C (non réparé par le MMR).

La protéine MutL n'est pas essentielle à la formation de ces boucles puisque les boucles α sont également observées en présence de MutS et d'ATP seuls. Mais la présence de MutL permet d'augmenter la vitesse de formation de ces boucles d'un facteur 1.7 (de 380bp/min à 630bp/min). Ces deux protéines, révélées par l'utilisation d'anticorps anti-MutS et anti-MutL, sont sous forme dimérique et se situent à proximité l'une de l'autre, à la base de ces boucles.

Enfin, l'absence de boucles en présence d'ATP γ S non-hydrolysable confirme l'importance de l'hydrolyse de l'ATP pour la formation de ces boucles. L'hydrolyse de l'ATP semble fournir l'énergie pour la translocation et la formation d'une boucle α serait un moyen simple pour permettre à la protéine de scanner l'ADN dans les deux sens afin de chercher le signal de discrimination des deux brins et déclencher la réparation.

- Le modèle de commutateur moléculaire (**Figure 1 - 3 au milieu**)

Ce modèle, proposé par R. Fishel, diffère de celui de translocation, essentiellement par le rôle joué par l'ATP. Comme précédemment, MutS lié à l'ATP est capable de diffuser le long du duplex d'ADN, en présence ou non de MutL. Mais, là où, dans le modèle précédent, l'hydrolyse de l'ATP joue le rôle de moteur moléculaire, dans le modèle de commutateur moléculaire, l'hydrolyse de l'ATP permet simplement au complexe MutS*MutL de se détacher de l'ADN (Acharya et al., 2003). En effet, les expériences de retard sur gel semble identique que l'on ajoute MutS, MutL et de l'ATP ou MutS en l'absence d'ATP; comme MutS en l'absence d'ATP se fixe sur le mésappariement avec une bonne affinité, la présence de MutL et d'ATP se traduit par l'hydrolyse de l'ATP fixée sur MutS (la fixation d'ATP sur MutL n'est requise que pour l'activation de MutH). Suite à l'hydrolyse d'ATP, MutS se retrouve lié à l'ADP et dans ces conditions, semble être recyclée pour lier un

nouveau mésappariement. Conformément à ce modèle, si l'on remplace l'ATP par de l'ATP γ S non-hydrolysable, le complexe MutS*MutL reste lié de fortement à l'ADN.

Dans ce modèle, une fois que MutS a recruté MutL, le complexe diffuse en 5' ou 3' à plus d'un 1kb du mésappariement jusqu'au site hémiméthylé, l'hydrolyse de l'ATP libérant ce complexe suite à l'activation de MutH. Le fait que MutS*MutL diffuse le long de l'ADN laisse le mésappariement libre pour qu'un autre dimère de MutS vienne s'y fixer puis après échange ADP \rightarrow ATP, diffuse à son tour vers un site hémiméthylé. Il est alors probable selon ce modèle que la reconnaissance et la diffusion soient des phénomènes itératifs. Plusieurs complexes MutS*MutL sont recrutés et diffusent les uns derrière les autres vers un site de discrimination. Le phénomène itératif ne cesserait que lors de l'excision du brin entre le site hémiméthylé et le mésappariement.

– Le modèle stationnaire (**Figure 1 - 3 à droite**)

Dans ce modèle proposé par P. Hsieh et W. Yang, l'interaction de MutL avec le complexe MutS*ADN*ATP arrête MutS dans sa diffusion le long de l'homoduplex d'ADN et l'oblige à rester à proximité du mésappariement (Schofield et al., 2001). Sans quitter la proximité du mésappariement, MutS*MutL active alors l'endonucléase MutH *via* l'interaction entre MutL et MutH. Il se forme donc une boucle où l'ADN se retrouve pincé entre le site du mésappariement et le site hémiméthylé et à la base de laquelle se trouvent MutS et MutL. La boucle est donc similaire à celle décrite dans le modèle de translocation, à la différence que le mésappariement se localise à la base de la boucle. Des boucles où MutS*MutL et le mésappariement sont au voisinage de la base, sont également observées en microscopie électronique (Allen et al., 1997). En accord avec ce modèle, il a été observé *in vitro* que le complexe MutS*MutL active MutH en présence d'ATP et de Mg²⁺, dans une réaction en *trans* où le mésappariement et le site hémiméthylé sont portés par deux molécules d'ADN séparées (Junop et al., 2001). L'activation de MutH est, par ailleurs, aussi efficace en *trans* que lorsque les deux sites sont portés par la même molécule d'ADN. Le complexe MutS*MutL ne semble donc pas avoir besoin de diffuser le long de l'ADN jusqu'au site hémiméthylé pour activer MutH.

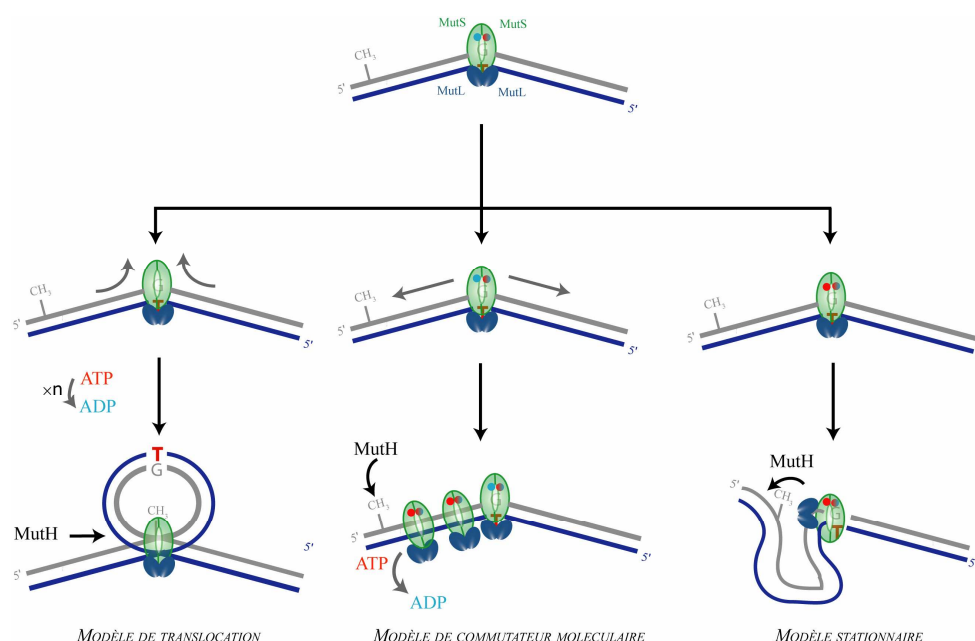


Figure 1 - 3: Schéma des trois modèles de communication entre le mésappariement et l'incision au niveau du site d'hémiméthylation (d'après (Li, 2008))

Ce dernier modèle est contredit par le fait que l'activation de MutH est inhibée *in vitro* lorsqu'on insert un site de reconnaissance de l'endonucléase EcoRI entre le mésappariement et le site hémiméthylé, en présence de l'enzyme inactivée EcoRI_{E111Q} (la mutation E111Q implique que EcoRI se fixe sur son site de reconnaissance sans l'inciser) (Pluciennik and Modrich, 2007); dans cette expérience, l'enzyme de restriction forme un bloc empêchant la diffusion entre les deux sites. Dans ces conditions, MutH n'est plus activé et il semble donc que le complexe MutS*MutL ait besoin de diffuser jusqu'au site d'hémiméthylation pour activer MutH. Ces expériences confirment les modèles où le complexe MutS*MutL diffuse vers le site de discrimination.

De plus, le taux d'hydrolyse d'ATP par MutS est relativement faible, de l'ordre de 1 à 10min⁻¹ suivant les conditions expérimentales (Acharya et al., 2003; Bjornson et al., 2000). Cette vitesse d'hydrolyse est difficilement conciliable avec le modèle de translocation. Le modèle favorisé à ce jour est donc celui de commutateur moléculaire même si on ne peut exclure que les protéines MutS, MutL et MutH utilise des mécanismes empruntant un peu à chacun de ces modèles.

Réparation du mésappariement

Seule, MutH possède une activité endonucléase faible et est spécifique aux motifs GATC non-méthylés (Welsh et al., 1987). Un crible de double-hybride a mis en évidence que MutL interagit avec MutH alors que MutS ne semble pas faire d'interaction, indiquant que MutL joue probablement le rôle d'activateur. Ainsi l'ajout de MutL en présence d'ATP (ou d'analogues non-hydrolysables) stimule

l'activité endonucléase de MutH, et ce, en l'absence de MutS; MutS ne stimule l'activité enzymatique qu'en présence d'un mésappariement (Hall and Matson, 1999).

Une fois activée par MutL, MutH clive le brin néosynthétisé en 5' de l'adénine non-méthylée. L'endonucléase MutH est fonctionnelle sous forme monomérique et contient dans son site actif, deux cations (Ca^{2+} dans le cas de la structure de MutH de *H. influenzae*, hiMutH (Lee et al., 2005)). Les structures cristallographiques de MutH d'*E. Coli* (Ban and Yang, 1998b) et de MutH de *H. influenzae* montrent que la protéine est composée de deux bras (bras N-terminal et C-terminal) reliés par une région flexible (**Figure 1 - 4(A)&(B)**). Il a été proposé que le mouvement de ces deux bras joue le rôle d'interrupteur régulant l'activité endonucléase de MutH *via* l'interaction entre MutL et l'hélice α_E de MutH (Giron-Monzon et al., 2004).

Il reste à déterminer comment MutH discrimine le brin méthylé du brin non-méthylé. En analysant les structures de hiMutH complexé avec une molécule d'ADN non-méthylée et hémiméthylée, le groupe méthyle porté par l'adénine est spécifiquement reconnu par la proline P185; cette reconnaissance a deux conséquences: d'une part, la position de la proline sur MutH implique que c'est forcément le brin non-méthylé qui se fixe dans le site catalytique; d'autre part, la reconnaissance par la proline P185 semble induire un réarrangement conformationnel sur toute la structure, qui se traduit par un rapprochement des résidus du site catalytique (E56, D70, E77 et K79). Ce rapprochement permet alors à ces résidus de s'aligner avec les deux cations Ca^{2+} et rend ainsi possible l'attaque nucléophile pour cliver le brin d'ADN (**Figure 1 - 4(C)**).

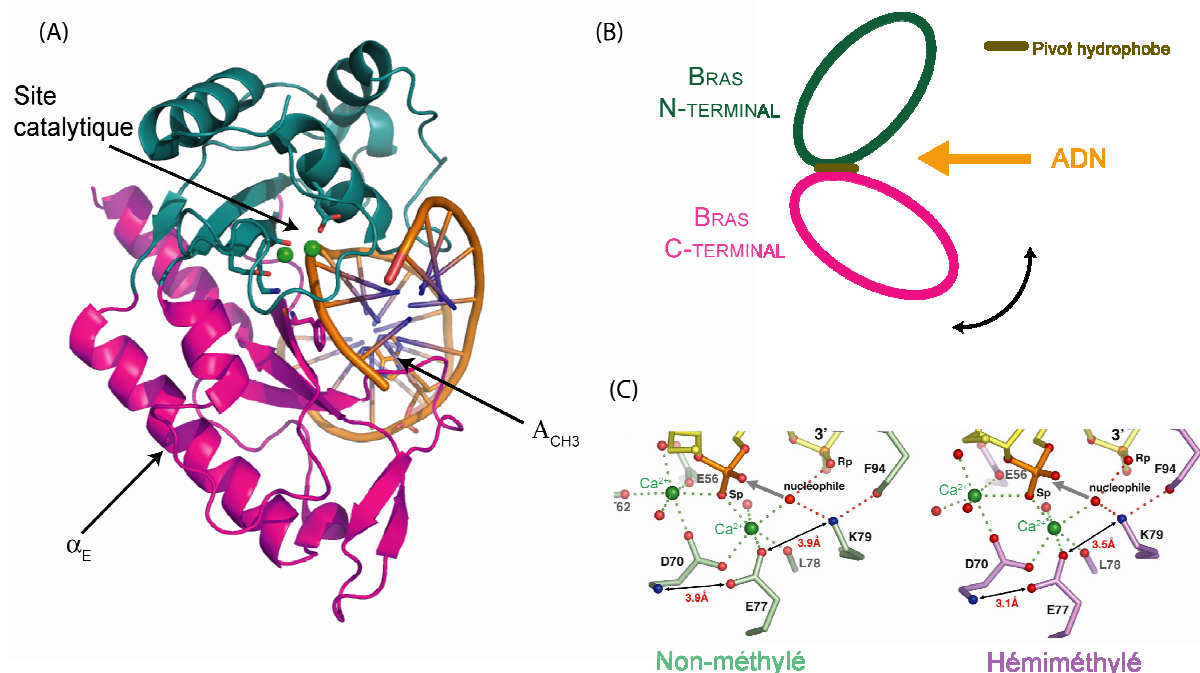


Figure 1 - 4: Structure de l'endonuclease MutH (A) Structure cristallographique (code PDB:2AOQ) à 2.2Å de hiMutH lié à une double hélice d'ADN non-méthylé (en orange); les bras N- et C-terminaux sont respectivement colorés en vert marin et rose; les cations Ca^{2+} sont représentés par des sphères vertes (B) Schéma simplifié de l'organisation de MutH et du mouvement de pivot du bras C-terminal (en rose) par rapport au bras N-terminal (en vert) (C) Comparaison des configurations du site actif de hiMutH en complexe avec l'ADN non-méthylé (à gauche) et hémiméthylé (à droite); la coordination des cations est représentée en pointillés verts et les liaisons hydrogène en pointillés rouges; les distances représentées sont celles subissant les plus grands changements entre les deux configurations (d'après (Lee et al., 2005))

L'incision créée par MutH semble être le signal d'entrée pour l'hélicase II (ou UvrD). UvrD seul ne transloque que sur une distance de 40 à 250bp (Ali and Lohman, 1997; Dessinges et al., 2004). Ceci est insuffisant pour joindre le mésappariement au site hémiméthylé qui peut se trouver à une distance supérieur à 1kb. Il a ainsi été montré que UvrD interagit avec MutL en double-hybride (Hall et al., 1998) et que la présence de MutL stimule l'activité hélicase de UvrD (Yamaguchi et al., 1998). Cette interaction semble avoir trois actions sur UvrD: (i) d'une part, il a été observé par des expériences de retard sur gel que la présence de MutL stabilise l'interaction entre UvrD et l'ADNsb (Mechanic et al., 2000). (ii) D'autre part, le mutant MutL-R226E qui présente une faible affinité pour l'ADN par rapport à la protéine sauvage, réduit considérablement l'activation de l'hélicase (Robertson et al., 2006). De ce fait, il en a été déduit que MutL est important pour orienter UvrD vers l'un ou l'autre des brins de l'ADN; ainsi, comme UvrD n'est active que dans le sens 3'→5' et que le site de discrimination peut se trouver en 5' ou en 3' du mésappariement, l'interaction entre MutL et UvrD semble permettre à MutL d'indiquer à UvrD le brin sur lequel se fixer pour assurer que le MMR soit bidirectionnel et que l'excision se fasse bien entre le site hémiméthylé et le mésappariement. (iii) Enfin le dernier rôle de cette interaction MutL*UvrD est d'augmenter la distance de translocation de l'ADN; un mécanisme

faisant intervenir un processus itératif a d'ailleurs été proposé : une première hélicase déroule la double hélice d'ADN selon sa processivité intrinsèque; MutL recrute alors une seconde hélicase qui, selon sa propre processivité, va dérouler le brin d'ADN en commençant là où la première hélicase s'est arrêtée. Ce mécanisme se réitère jusqu'à ce que l'ensemble de la séquence entre le site hémiméthylé et le mésappariement soit sous simple brin (Matson and Robertson, 2006).

L'ADN déroulé est alors stabilisé par les protéines Ssb (Single-strand binding protein) avant que le brin d'ADN néosynthétisé ne soit dégradé dans le sens 5'→3' par les exonucléases ExoVII et RecJ et dans le sens 3'→5' par les exonucléases ExoI, ExoVII et ExoX (ExoVII possédant une double activité exonucléasique) (Burdett et al., 2001).

L'ADN polymérase III est recrutée probablement *via* son interaction avec les protéines MutS et MutL et resynthétise le brin d'ADN (Pluciennik et al., 2009). Les extrémités d'ADN sont alors jointes par l'ADN ligase.

En conclusion, la protéine MutL ne présente pas d'activité catalytique propre mais joue chez *E. coli* le rôle de chef d'orchestre, coordonnant les différentes étapes en aval de la reconnaissance du mésappariement. Cette protéine intervient ainsi dans de nombreuses interactions directes avec les différents partenaires pour inciser, dérouler, dégrader et resynthétiser le brin pour éliminer l'erreur d'appariement. Les mécanismes moléculaires qui permettent à MutL d'effectuer de façon coordonnée l'ensemble de ces activités vis-à-vis de MutS, MutH, UvrD et Pol III ne sont cependant pas complètement élucidés.

iii. Le mécanisme MMR post-répliatif chez les eucaryotes

Le mécanisme de réparation des mésappariements est assez bien conservé chez les eucaryotes et a beaucoup de points communs avec le MMR des procaryotes. Cependant il existe certaines différences au niveau des protéines et du mécanisme détaillé ci-dessous ([Tableau 1 - 1](#)).

Tableau 1 - 1: Les protéines du MMR et leur fonction (modifié d'après (Kunkel and Erie, 2005))

<i>E. coli</i>	<i>S. cerevisiae</i>	<i>H. sapiens</i>	Fonction
MutS	scMutS α (Msh2*Msh6)	hsMutS α (MSH2*MSH6)	Reconnaissance des mésappariements et des IDLs de 1-2 bases
	scMutS β (Msh2*Msh3)	hsMutS β (MSH2*MSH3)	Reconnaissance des IDLs plus longs
MutL	scMutL α (Mlh1*Pms1)	hsMutL α (MLH1*PMS2)	Coordinateur moléculaire; activité endonucléase. Participe à la recombinaison méiotique
	scMutL β (Mlh1*Mlh2)	hsMutL β (MLH1*PMS1)	Fonction inconnue chez l'humain; participerait à la réparation d'IDLs chez la levure
	scMutL γ (Mlh1*Mlh3)	hsMutL γ (MLH1*MLH3)	Participe à la recombinaison méiotique et à la réparation d'IDLs; activité endonucléase
MutH	<i>Non identifié</i>	<i>Non identifié</i>	Endonucléase; incise le brin néosynthétisé au niveau du site GATC hémiméthylé
UvrD	<i>Non identifié</i>	<i>Non identifié</i>	Hélicase; déroule l'ADN après incision
RecJ ExoVII	<i>Non identifié</i>	<i>Non identifié</i>	Exonucléases 5'→ 3'; clive le brin néosynthétisé après le passage d'UvrD
ExoI ExoVII ExoX	<i>Non identifié</i>	<i>Non identifié</i>	Exonucléases 3'→ 5'; clive le brin néosynthétisé après le passage d'UvrD
<i>Non identifié</i>	Exo1	EXO1	Exonucléase 5'→ 3' agissant de préférence sur l'ADN double brin
Ssb	RPA	RPA	Protéine fixant l'ADN simple brin et assurant la protection de l'ADN simple brin, le temps de l'étape d'excision et de resynthèse
$\gamma\delta$	RFC	RFC	Facteur de chargement recrutant l'anneau β (<i>E. coli</i>) ou PCNA (eucaryotes) sur l'ADN
β	PCNA	PCNA	Facteur de processivité améliorant l'efficacité de la polymérase; interagit avec un grand nombre de protéines impliquées dans la réparation et/ou la recombinaison de l'ADN
ADN pol III	ADN pol δ	ADN pol δ	Synthèse de l'ADN
ADN ligase	ADN ligase I	ADN ligase I	Ligation après la synthèse de l'ADN

Reconnaissance du mésappariement sur le brin d'ADN

Chez les eucaryotes, plusieurs homologues des protéines MutS ont été identifiés. Six homologues ont notamment été trouvés chez la levure et chez l'homme, nommés Msh1 à 6 pour MutS homolog. Les protéines Msh forment des hétérodimères à la différence de l'homodimère MutS trouvé chez les bactéries. Des études de retard sur gel ainsi que des études génétiques ont montré que les mésappariements base/base et les boucles d'insertion/délétion de 1 à 2 nucléotides sont reconnus par l'hétérodimère Msh2*Msh6 (appelé également MutS α) (Alani et al., 1997; Marsischky et al., 1996; Marsischky and Kolodner, 1999). La reconnaissance des IDLs contenant jusqu'à 12 nucléotides est effectuée par un second hétérodimère Msh2*Msh3 (appelé également

MutS β) (Habracken et al., 1996). Le complexe Msh2*Msh3 de levure est donc partiellement redondant avec Msh2-Msh6 car il intervient dans la réparation des IDLs de 2 à 8 nucléotides.

Les données concernant des souris "knock out" chez lesquelles ont été invalidés un ou plusieurs gènes, viennent confirmer les résultats obtenus chez la levure (Wei et al., 2002). Les souris *msh2*^{-/-} sont viables et fertiles mais présentent une espérance de vie très réduite (50% meurent avant 6 mois). La cause principale de mortalité est le développement de lymphomes. Les extraits cellulaires de fibroblastes de ces souris n'ont pas d'activité de réparation des mésappariements de type base/base et des IDLs de plus de 4 nucléotides. Les souris double-mutantes *msh6*^{-/-} et *msh3*^{-/-}, possèdent un phénotype similaire aux souris *msh2*^{-/-}.

Les souris *msh6*^{-/-}, également viables et fertiles, ont une espérance de vie réduite due à des tumeurs, mais dans ce cas, il faut attendre 8-10 mois pour que 50% des souris "KO" meurent. L'analyse d'extraits cellulaires de ces souris met en évidence une déficience dans la réparation des mésappariements base/base et une conservation de la réparation des IDLs.

Enfin, les souris *msh3*^{-/-} ne présentent pas de comportement les distinguant des souris sauvages. En effet, ces souris réparent correctement les mésappariements base/base et les IDLs de taille supérieure à deux nucléotides. Elles présentent, en revanche, un défaut de réparation des IDLS courts de deux nucléotides.

Trois structures cristallographiques de l'hétérodimère humain MutS α (MSH2*MSH6) en complexe avec différents duplexes d'ADN ont été résolues par J.J. Warren et al. (Warren et al., 2007). Les protéines utilisées en cristallogénèse correspondent à la protéine MSH2 entière ainsi qu'à un fragment de MSH6 ne contenant pas les 340 premiers résidus de l'extrémité N-terminale. Ce complexe a une activité de reconnaissance des mésappariements proche de la protéine sauvage *in vitro*. Deux duplexes d'ADN ont été utilisés contenant soit un mésappariement G.T soit une insertion d'une thymine. La structure globale du complexe est proche de celle de MutS d'*E. coli* en complexe avec un duplex d'ADN (**Figure 1 - 5(A)**). Chaque monomère possède trois domaines fonctionnels: un site de reconnaissance du mésappariement, un site «pince» et un site ATPase. La reconnaissance de l'erreur d'appariement s'effectue de manière asymétrique; seul le site de reconnaissance de MSH6 se lie spécifiquement à la thymine mésappariée *via* la phénylalanine F432 du motif conservé FxE (Dufner et al., 2000). Le site de reconnaissance sur MSH2 ainsi que les domaines «pince» des deux monomères se lie de manière non-spécifique à l'ADN (**Figure 1 - 5(B)**). Cette reconnaissance du mésappariement implique une torsion de l'ADN d'environ 45°, similaire à ce qui est observé chez MutS.

De plus, nous avons indiqué précédemment que l'homodimère de MutS n'avait pas la même affinité sur ses deux sites ATPases. La même conclusion a été observée

lors de mesures d'interaction de MutS α avec des nucléotides par anisotropie de fluorescence (Martik et al., 2004). La structure cristallographique de MutS α obtenue sans ajout d'ADP à la solution de cristallisation, montre que MSH2 contient une molécule d'ADP au sein de son site ATPase, le site ATPase de MSH6 étant vide.

La structure d'une partie des 340 résidus de MSH6 non cristallisés, *i.e.* entre les résidus 68 et 201, a été résolue par RMN au LBSR (Laguri et al., 2008). Ce domaine possède un repliement de type PWWP se liant à l'AND double-brin. Il a été proposé que ce domaine stabilise l'interaction entre MutS α et l'ADN.

Les structures cristallographiques de MutS α montrent que le comportement de MutS α vis-à-vis de la liaison à l'ADN ou de la fixation de nucléotides, est comparable à celui de MutS chez les bactéries. Il est donc probable, et les données biochimiques *in vitro* semblent le confirmer, que MutS α reconnaisse l'erreur d'appariement selon un mécanisme proche de celui de MutS.

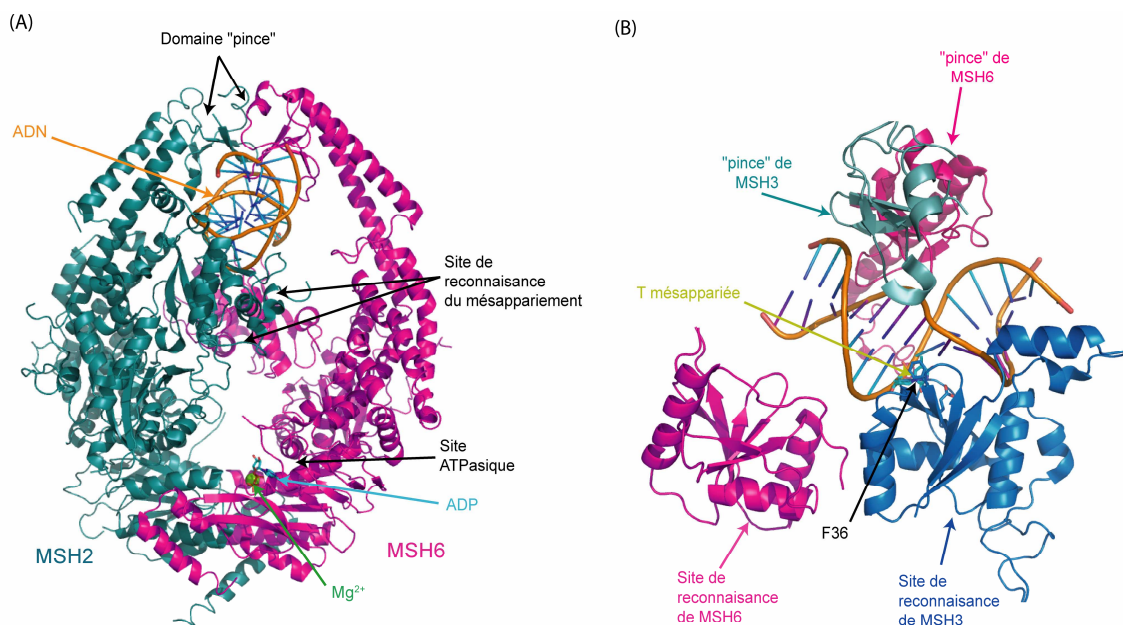


Figure 1 - 5 : Structure du complexe hMutS α (A) Structure cristallographique (code PDB:2O8E) à 3.3Å de hMutS α lié à une double hélice d'ADN portant un mésappariement G-T (en orange); MSH2 et MSH6 sont respectivement colorés en vert marin et magenta; le cation Mg²⁺ est représenté par une sphère verte (B) Détail du site de fixation de hMutS α à une thymine mésappariée; le site de reconnaissance se liant spécifiquement à la thymine est représenté en magenta, le second site est représenté en bleu; les deux domaines pincés sont représentés en rose et vert marin

Il n'existe pas de structure cristallographique du complexe MutS β (MSH2*MSH3). La modélisation par homologie à partir de la structure de MutS α montre que la phénylalanine conservée de MSH6 est remplacée par une arginine conservée K158 sur MSH3; il a été proposé que la présence de cet acide aminé, K158 dans MSH3 serait un des éléments importants dans la spécificité de MutS β par rapport à MutS α . La position K158 créerait un encombrement stérique incompatible avec le mode de reconnaissance d'un mésappariement base/base observé pour MutS α .

(Downen et al., 2010). Cette observation est en accord avec une étude au cours de laquelle une protéine chimère a été construite avec la séquence globale de Msh6 de *S. cerevisiae* et le domaine de reconnaissance de Msh3. Dans cette chimère, les 116 résidus du domaine de reconnaissance de Msh6 ont été remplacés par les 130 résidus du domaine équivalent de Msh3 (Shell et al., 2007). Cette chimère reconnaît spécifiquement les IDLs longs de 3-4 nucléotides selon des mesures par résonance plasmonique de surface, confirmant que le domaine de reconnaissance du mésappariement est important pour la spécificité de substrat de chaque hétérodimère.

De la même façon, les eucaryotes présentent plusieurs protéines homologues des protéines MutL bactériennes. On dénombre quatre homologues chez la levure et chez l'homme; les protéines Mlh1 à 3 pour MutL homolog et Pms1 pour Post meiotic segregation 1 chez la levure; les protéines Mlh1, Mlh3, Pms1 et Pms2 chez l'homme. Ces homologues forment trois hétérodimères chez la levure: MutL α (Mlh1*Pms1), MutL β (Mlh1*Mlh2) et MutL γ (Mlh1*Mlh3) et trois hétérodimères MutL α (MLH1*PMS2), MutL β (MLH1*PMS1) et MutL γ (MLH1*MLH3) chez l'homme. Il est à noter que la protéine humaine PMS2 est l'homologue la plus proche de la protéine Pms1 de levure. La protéine PMS1 chez l'homme correspond à la protéine Mlh2 de *S. cerevisiae*. La voie de réparation des mésappariements implique essentiellement l'hétérodimère MutL α . Les autres hétérodimères sont impliqués dans d'autres processus cellulaires comme la recombinaison méiotique pour l'hétérodimère MutL γ . Les données fonctionnelles et structurales de ces différents complexes seront reprises plus en détail § 1.1.2.

Discrimination du brin néosynthétisé portant le mésappariement

Les mécanismes moléculaires qui permettent la discrimination des brins néosynthétisé et parental ne sont pas totalement résolus chez les eucaryotes. Aucune protéine homologue de l'endonuclease MutH n'a été identifiée chez les eucaryotes. Des analyses *in vitro* avec des systèmes MMR de levure ou humain reconstitués, la réparation des mésappariements dépend de la présence d'une incision en 5' ou 3' sur l'un des brins, de la présence de PCNA et de la protéine qui charge PCNA, RFC (Dzantiev et al., 2004; Thomas et al., 1991). Les essais développés au sein du laboratoire de P. Modrich, utilisent un plasmide contenant un mésappariement et une incision en 5' ou en 3' du mésappariement ainsi que les protéines purifiées suivantes: MutS α , MutL α , Exo1, PCNA et RFC. Ils ont montré qu'en l'absence d'incision en 5' ou 3', la réparation du mésappariement n'est pas observée. Or, après le passage de la fourche de réplication, le brin néosynthétisé, portant donc le mésappariement, contient plusieurs incisions, que cela soit au niveau des extrémités 3' ou des extrémités 5' des fragments d'Okazaki. Ces terminaisons 5' ou 3' ont été proposées comme servant de signal principal pour la discrimination des deux brins.

Ceci semble confirmer que le mécanisme MMR a une "relation particulière avec le complexe de réplication" (Wagner and Meselson, 1976), et notamment avec le trimère PCNA (ou Pol30 chez *S. cerevisiae*, équivalent de la pince β , ou " β clamp" en anglais, des bactéries).

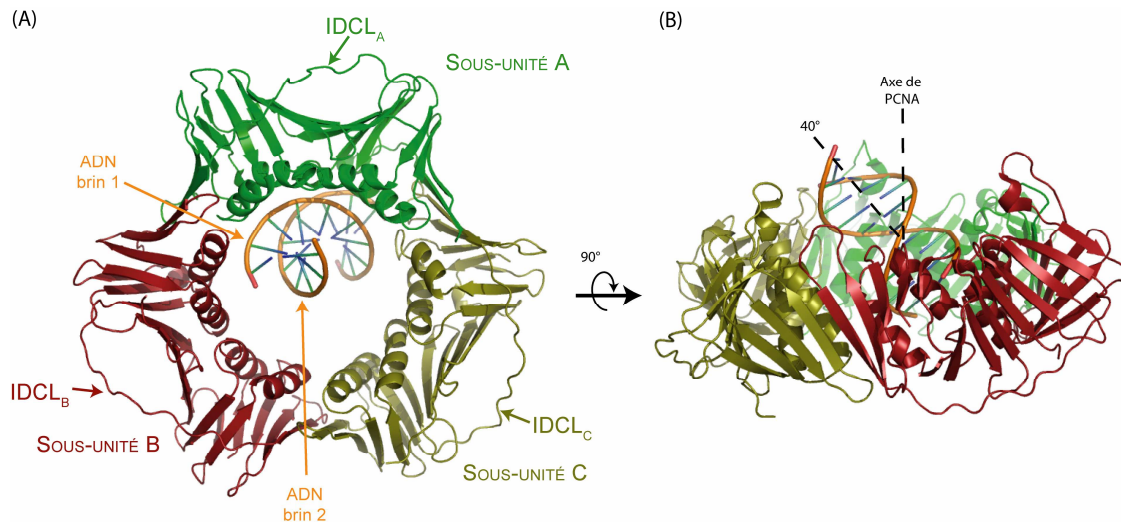
Des interactions protéine-protéine ont été observées entre la protéine PCNA et différents homologues MutS et MutL. L'analyse des séquences des protéines Msh a mis en évidence la présence d'un motif d'interaction avec PCNA, appelé motif PIP (ou PIP-box PCNA interacting protein box en anglais). Ce motif est situé en région N-terminale des protéines Msh3 et Msh6. Le motif PIP est constitué de la séquence Qxx[L/M/I]xxF[F/Y] (Moldovan 2006 Cell) et est situé en région C-terminale ou N-terminale des protéines qui le contiennent. Des études structurales ont montré que ce motif adopte une structure en brin β suivie d'une hélice 3_{10} et interagit avec une région de PCNA appelée IDCL pour Interdomain connecting loop. Le trimère de PCNA présente trois sites potentiels de fixation d'un motif PIP. Il a été montré que le complexe Msh2-Msh6 de levure co-immunoprécipite (co-IP) avec PCNA et que la présence de PCNA favorise dans des expériences de retard sur gel, la fixation du complexe MutS α sur un mésappariement (Flores-Rozas et al., 2000). En revanche, le variant Msh2*(Msh6-FF33AA) muté sur les phénylalanines faisant partie du motif PIP (₂₇QSSLLSFF) de Msh6, n'interagit plus dans des essais de co-IP avec PCNA. De même, les protéines humaines MSH3 et MSH6 interagissent avec PCNA par co-IP *via* leur motif PIP présent dans leur domaine N-terminal (Kleczkowska et al., 2001) mais à la différence de la levure, PCNA n'a pas d'effet sur la fixation de MutS α humain sur un mésappariement d'après des mesures par résonance plasmonique de surface (Iyer et al., 2008).

Lors d'un crible de double hybride réalisé sur les protéines entières Mlh1, Pms1, PMS2 comme appât, un fragment cDNA correspondant aux protéines PCNA de levure et humaine ont été isolés. Le signal observé en termes d'activité β -galactosidase pour ces deux interactions est plus faible que celui mesuré dans la même étude pour la formation des hétérodimères Mlh1/Pms1 (Umar 1996 Cell). Un crible double-hybride a été réalisé dans la même étude en utilisant comme appât les complexes LexA-Mlh1/Pms1 ou Mlh1/LexA-Pms1 de levure. La protéine PCNA de levure a également été retrouvée comme partenaire. Les auteurs ont ensuite mis en évidence, un mutant de PCNA (P115S) qui augmente d'un facteur 150 chez la levure, le taux de mutation sur une séquence répétée poly-GT de 50 paires de bases. Ce taux de mutation est proche de celui d'une souche *mlh1 Δ* (augmentation 190 fois) et du double mutant *pcna-P115S/mlh1 Δ* (augmentation 180 fois). Plus récemment, une étude a proposé la présence d'un motif PIP dégénéré (QxxLxxFA) sur la région C-terminale de Mlh1 de levure (Lee and Alani, 2006). Cette observation s'appuie sur des mesures par résonance plasmonique de surface. De façon surprenante et contraire aux autres motifs PIP étudiés, ce motif est toujours fonctionnel si un des trois acides aminés Q,L ou F est muté en alanine. Seule la triple mutation du motif QLF annule

l'interaction et s'accompagne d'un phénotype mutateur d'un facteur 4900 dans un essai *Lys2::A14* ; ce facteur est proche d'une délétion *mlh1Δ* (facteur 5000).

Le fait que PCNA interagisse avec les protéines intervenant dans les premières étapes du MMR, suggère que PCNA est impliqué dans des étapes antérieures à celle de resynthèse. Umar et al. ont constitué un système à partir d'extraits cellulaires où la brèche générée au cours de l'étape d'excision du MMR peut être remplie par une polymérase d'*E. coli*. Ainsi, ces extraits cellulaires en présence d'un plasmide contenant une incision et un mésappariement sont fonctionnels jusqu'à l'excision du brin; L'ajout d'un peptide de p21, un des motifs PIP les plus affins de PCNA, bloque les premières étapes du MMR, indiquant que PCNA est requis dès les premières étapes du MMR (Umar et al., 1996). Le rôle de PCNA n'est pas totalement élucidé. Il est cependant avancé qu'un de ses rôles serait de faire le lien entre le complexe MutS*MutL eucaryote lié au mésappariement et le signal de discrimination, à savoir une incision en 5' ou 3' du mésappariement. En effet, il a été observé par microscopie électronique que PCNA pouvait se fixer à une extrémité 3' d'un ADN circulaire contenant une incision, et ce, par la seule présence de RFC qui joue le rôle de protéine de chargement et d'ATP. Aucune polymérase n'est donc nécessaire, montrant que PCNA peut se fixer sur une extrémité lors des premières étapes du MMR (Podust et al., 1995). La protéine PCNA est un trimère formant un anneau présentant deux faces non-équivalentes, l'une tournée vers la polymérase lors de la réplication, l'autre vers les protéines partenaires de PCNA dont les protéines du MMR précitées (Krishna et al., 1994) et récemment, une structure cristallographique a été résolue à 3Å entre le trimère de *S. cerevisiae* et un ADN comportant une extrémité simple brin de 5 nucléotides (**Figure 1 - 6(A)**). Cette structure a permis d'avoir un modèle de la liaison de PCNA à l'ADN (McNally et al., 2010): PCNA se lie spécifiquement à l'ADN et de manière asymétrique puisqu'un des deux brins (le brin 1) forme une grande partie des contacts avec PCNA; les contacts spécifiques impliquent que l'ADN n'est pas perpendiculaire à l'anneau PCNA mais forme un angle de 40° avec l'axe de symétrie du trimère (**Figure 1 - 6(B)**). Du fait d'une résolution trop faible et du mouvement opéré par une partie du brin d'ADN cristallisé, il n'est pas possible de voir comment se fixe l'extrémité simple brin. Mais les similitudes entre la structure PCNA*ADN et celle du β-clamp d'*E. coli* lié également à un ADN possédant une extrémité simple brin (Georgescu et al., 2008), permettent de proposer que comme chez *E. coli*, l'ADN simple brin vient se fixer spécifiquement dans une cavité hydrophobe de PCNA.

Le fait que PCNA puisse se lier spécifiquement à une incision du brin néosynthétisé, laisse supposer que le trimère de PCNA oriente, de par son positionnement asymétrique sur l'ADN, le complexe MMR vers le brin néosynthétisé.



Réparation du mésappariement

La reconstitution *in vitro* de systèmes de réparation MMR eucaryotes par le laboratoire de P. Modrich a été une étape importante dans la compréhension du rôle des différents facteurs impliqués dans l'étape de réparation du MMR (Dzantiev et al., 2004; Genschel and Modrich, 2003) (Figure 1 - 7). Ces travaux ont montré que dans le cas d'une réparation dirigée en 5' du mésappariement, la réparation a lieu en présence d'un mésappariement, d'une incision, de MutS α , Exo1, RPA et d'ATP. Dans le cas d'une réparation dirigée en 3' du mésappariement, la réparation nécessite la présence de MutL α , de PCNA et de RFC en plus des facteurs précités. Ce système permet également de réparer une insertion de trois nucléotides en remplaçant MutS α par MutS β (Zhang et al., 2005).

A la différence d'*E. coli*, Exo1 est la seule activité exonucléase associée au MMR connue à ce jour. De façon intéressante, les levures *exo1Δ* (Tran et al., 2001) présentent une faible augmentation du taux de mutation dans un essai spécifique du MMR, *i.e.* le taux de mutation sur l'allèle *hom3-10*. La délétion *exo1Δ* présente une augmentation du taux de mutation d'un facteur 8 comparé à une augmentation du taux de mutation observé chez les levures *mlh1Δ* d'un facteur 1118. Ce résultat suggère des voies alternatives d'excision en l'absence d'Exo1. Une étude récente sur des extraits cellulaires obtenus à partir de cellules de souris *exo1^{-/-}* (Kadyrov et al., 2009) suggère que le mésappariement pourrait être ôté par un mécanisme de déplacement de brins mettant en jeu la polymérase δ (ou strand displacement synthesis en anglais). Des mesures des instabilités microsatellites ont été effectuées sur des modèles animaux déficients en protéines Exo1 ont été construits (Wei et al., 2003). Ces travaux montrent que les extraits cellulaires correspondant aux cellules

exo1^{-/-} présentent un défaut de réparation des répétitions de mono-nucléotides de type (A)_n comparés aux cellules de souris sauvages. Le défaut s'élève à 14% d'allèles instables pour les extraits *exo1*^{-/-} comparés à 3% pour des cellules de souris sauvages. La réparation des marqueurs di-nucléotides de type (CA)_n n'est pas affectée. En comparaison, dans des cellules *mlh1*^{-/-}, la réparation des deux types de marqueur, mono- et di-nucléotide, est affectée. L'analyse du phénotype mutateur des cellules *exo1*^{-/-} à un locus précis montre un taux de mutation 30 fois supérieur par rapport aux cellules sauvages et 5 fois inférieur à celui des cellules *msh2*^{-/-}. Ceci suggère que les cellules *exo1*^{-/-} gardent une certaine activité de réparation MMR. Les souris observées sur une période de 18 mois présentent une mortalité supérieure aux souris sauvages (50% de survie à 17 mois contre 85% pour les souris sauvages) mais inférieure aux souris *Msh2* (50% de survie à 8 mois). 70% des souris *exo1*^{-/-} qui atteignent l'âge de 16 mois développent des lymphomes. De façon intéressante, les souris *exo1*^{-/-} présentent de sévères défauts de méiose et sont stériles; ce point sera abordé un peu plus loin dans le cadre des autres rôles des protéines du MMR.

Exo1 fait partie de la famille Rad2 des exonucléases appelée «structure spécifique». Tous les homologues d'Exo1 (*S. cerevisiae*, *S. pombe*, *H.sapiens*) peuvent se fixer sur l'ADN double brin ou simple brin mais ont une préférence pour l'ADN double brin (Szankasi and Smith, 1992). Les protéines Exo1 possèdent une activité exonucléase 5'→3' sur l'ADN double brin (Qiu et al., 1999; Szankasi and Smith, 1995; Tishkoff et al., 1997). Ainsi, bien qu'unidirectionnelle, cette protéine est impliquée dans la réparation dirigée aussi bien en 5' qu'en 3' du mésappariement.

Lors d'une excision 5'→3', il a été montré *in vitro*, qu'EXO1 dégrade l'ADN entre l'incision et le mésappariement. L'activité d'EXO1 est largement favorisée en présence de MutSα, qui semble à la fois favoriser le recrutement d'EXO1 au niveau de l'incision et augmenter la capacité d'excision d'EXO1 (Genschel et al., 2002). Le rôle de MutSα a été confirmé par le fait que l'exonucléase interagit avec Msh2/MSH2 aussi bien chez l'homme (Rasmussen et al., 2000; Schmutte et al., 2001) que chez la levure (Tishkoff et al., 1997) par des essais de co-IP. Une interaction entre EXO1 et MSH3 a également été identifiée par des essais de capture (ou pull-down en anglais) (Schmutte et al., 2001). MutLα n'est pas essentiel pour l'excision 5'→3' mais sa présence augmente la sélectivité du système MMR pour les hétéroduplexes d'ADN (contenant un mésappariement) par rapport aux homoduplexes (Genschel and Modrich, 2003). De même, PCNA et RFC ne sont pas essentiels mais permettent d'augmenter l'efficacité de réparation d'environ 25% (Dzantiev et al., 2004).

Lors d'une excision 3'→5', MutLα devient essentiel puisqu'il incise le brin d'ADN néosynthétisé en 5' du mésappariement et permettre alors à Exo1 de dégrader dans le sens 5'→3' le brin néosynthétisé depuis l'incision faite par MutLα jusqu'après le mésappariement (**Figure 1 - 7**). Une interaction entre Exo1 et Mlh1 a été mise en évidence chez les eucaryotes par double-hybride et pull-down (Jager et al., 2001;

Schmutte et al., 2001). PCNA et RFC sont essentiels pour une excision 3'→5' car en plus du rôle de PCNA dans la discrimination des brins, PCNA stimule l'activité endonucléase de MutL α (Pluciennik et al., 2010). Comme il a été précisé auparavant, PCNA semble interagir avec Mlh1. il a été récemment montré au LBSR par calorimétrie qu'EXO1 possède un motif PIP₇₈₈QIKLNELW à l'extrémité C-terminale et que ce motif possède une affinité micromolaire pour PCNA (Liberti et al., 2010)

A la suite de l'excision du brin néosynthétisé au-delà du mésappariement, le brin est synthétisé une seconde fois par l'ADN polymérase Pol δ . La jonction entre le brin réparé et le reste du brin néosynthétisé est jointe par la ligase I. On peut noter qu'aucune activité hélicase n'a été associée à ce jour à la voie MMR eucaryote contrairement aux bactéries.

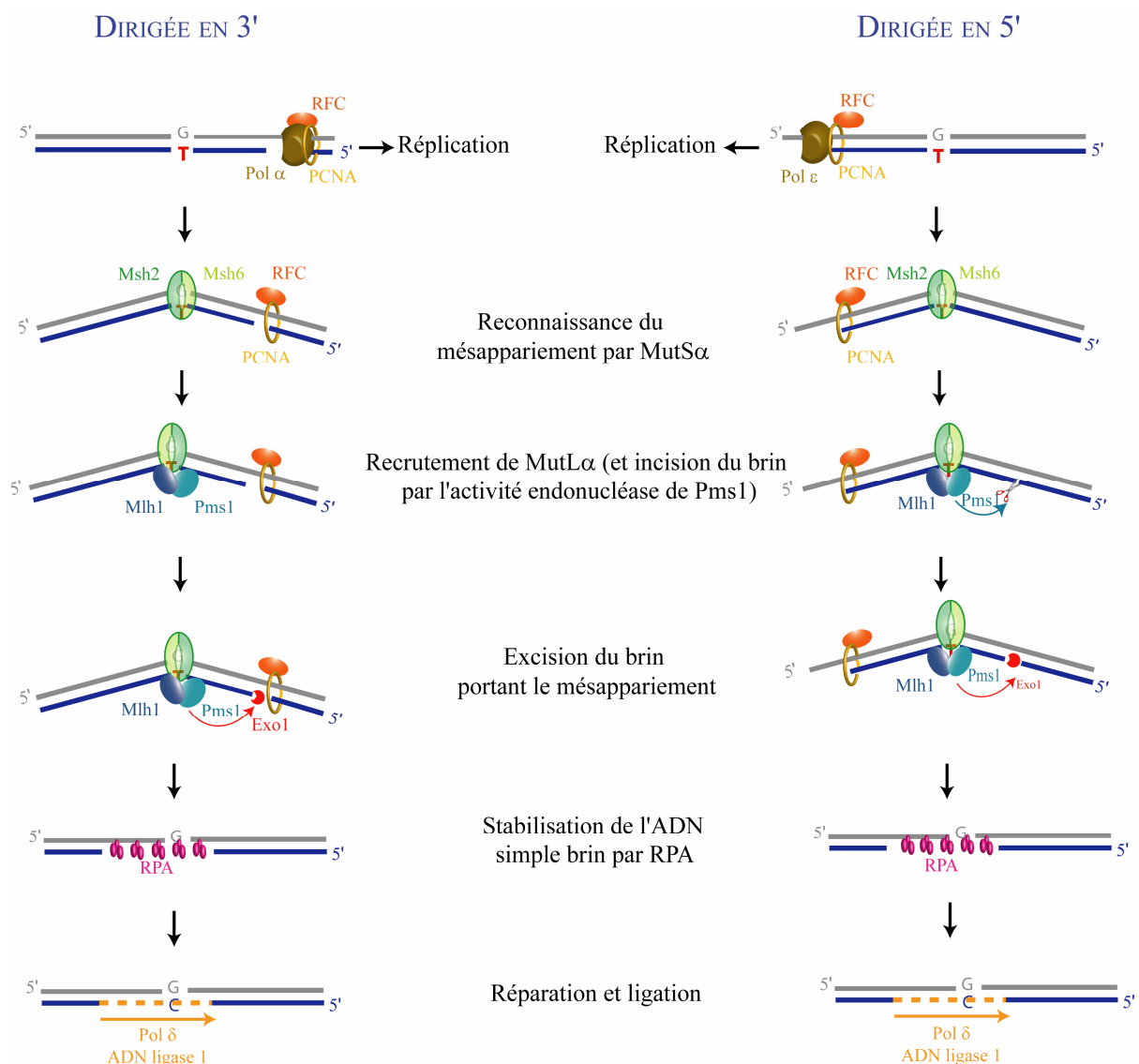


Figure 1 - 7: Schéma de la réparation d'un mésappariement G-T par le MMR chez les eucaryotes dans le cas où une incision se trouve en 3' du mésappariement (à gauche) ou en 5' du mésappariement (à droite)

iv. *Le mécanisme MMR chez les bactéries ne possédant pas de MutH*

Excepté pour les γ -protéobactéries proches d'*E. coli* qui dépendent de l'hémiméthylation, la majorité des eubactéries ne possèdent pas d'endonucléase MutH, semblant indiquer un mécanisme MMR similaire à celui des eucaryotes. De plus, chez certaines bactéries (*T. thermophilus*, *A. aeolicus* et *N. gonorrhoeae*), une activité endonucléase a été identifiée sur MutL. La similarité fonctionnelle des homologues de MutL de ces bactéries avec les homologues de PMS2 semble indiquer un mécanisme MMR de nature similaire, c'est-à-dire dépendant d'une incision comme signal de discrimination et non d'une endonucléase dirigée vers un site d'hémiméthylation.

Pour ces eubactéries, suite à la reconnaissance de l'erreur d'appariement, l'homodimère de MutS recrute MutL qui joue également le rôle de coordonnateur de la réparation du mésappariement. A l'instar du mécanisme eucaryote, l'existence d'une incision préexistante semble être le signal de discrimination et à travers l'interaction avec le β -clamp, va orienter le complexe MMR dans la réparation. Il a été montré que les deux protéines MutS et MutL interagissent avec le β -clamp chez *B. subtilis*. bsMutS interagit avec le β -clamp *via* un motif ₈₀₆QLSFF situé à son extrémité C-terminale (Simmons et al., 2008). La mutation simultanée de ces cinq résidus en alanine rend MutS incapable d'interagir avec le β -clamp et confère un phénotype mutateur 8 fois supérieur à la protéine sauvage (contre 60 fois pour une délétion *mutSΔ*). L'interaction entre bsMutL et le β -clamp a été confirmée par pontage (crosslinking en anglais) (Pillon et al., 2010b).

Jusqu'à très récemment, seules deux exonucléases, RecJ et ExoVII, avaient été identifiées chez les bactéries ne possédant pas MutH dans la voie MMR (Larrea et al., 2010). Ces exonucléases ayant une polarité 5'→3', cela suggérerait une excision du brin entre l'incision et le mésappariement identique à ce qui se passe avec Exo1 chez les eucaryotes. Cependant, il a récemment été découvert une exonucléase chez *T. thermophilus*, identifiée comme de la famille ExoI, car elle présente une activité 3'→5' spécifique à l'ADN simple brin (Shimada et al., 2010). De façon intéressante, la délétion de cette exonucléase augmente de 3 à 4 fois le taux de mutations chez *T. thermophilus*, par rapport à la souche sauvage. Le double mutant *exoIΔrecJΔ* provoque, pour sa part, une augmentation du taux de mutation d'environ 30 fois par rapport à la souche sauvage, alors que la délétion de RecJ présente un taux de mutations équivalent à une délétion d'ExoI. Ces résultats semblent indiquer que les deux exonucléases ExoI et RecJ sont toutes les deux impliquées dans le MMR mais à travers des voies différentes. Du fait de leur polarité inversée, il a été proposé que l'une ou l'autre interviennent suivant l'emplacement de l'incision, en 5' ou en 3' du mésappariement.

L'implication d'exonucléases spécifiques de l'ADN simple brin pose la question de l'implication d'hélicases dans le mécanisme de réparation. Un certain nombre d'hélicases homologues d'UvrD chez *E. coli* ont été identifiées, et notamment chez des organismes procaryotes ne présentant pas d'homologues de MutH. Cependant le

rôle activateur de MutL observé chez *E. coli* pour l'hélicase, ne semble pas être universel. Ainsi, chez trois bactéries, *T. tengcongensis*, *B. stearothermophilus* et *A. aeolicus*, la protéine MutL présente les motifs essentiels à l'activité endonucléase et l'activité endonucléase d'*A. aeolicus* a été caractérisée (Iino et al., 2010). Mais ces protéines n'ont pas le même comportement vis-à-vis des hélicases homologues d'UvrD identifiées dans ces organismes. *In vitro*, MutL n'est pas capable d'activer UvrD chez *T. tengcongensis* et *B. stearothermophilus*, (An et al., 2005). *A contrario*, MutL d'*A. aeolicus* peut activer un homologue d'UvrD, de manière similaire à l'activation d'UvrD chez *E. coli* (Mauris and Evans, 2010).

Les eubactéries autres que les γ -protéobactéries présentent donc un mécanisme MMR globalement plus proches de celui des eucaryotes que des bactéries proches d'*E. coli*. Chez les archaebactéries, le mécanisme de réparation MMR est beaucoup moins connu. Des homologues des protéines MutS et MutL ont été identifiés dans un nombre limité d'espèces. Il a été proposé que ces gènes provenaient d'un transfert horizontal de gènes depuis les bactéries (Larrea et al., 2010).

v. Les autres fonctions des protéines du MMR

Un certain nombre de protéines ont été identifiées comme interagissant avec les protéines Msh et Mlh. Ces travaux suggèrent que ces interactions correspondent à des fonctions des protéines Msh et Mlh dans voies cellulaires. L'étude de ces fonctions additionnelles est assez étonnante car elle décrit les protéines du MMR comme pouvant à la fois garantir la stabilité génomique mais également favoriser la variation génétique. De façon toute aussi paradoxale en première analyse, les protéines du MMR sont impliquées à la fois dans des mécanismes d'activation de l'apoptose et dans la survie cellulaire.

Le MMR comme médiateur de la réponse aux dommages de l'ADN

Il a été proposé que la voie de réparation MMR joue un rôle dans la détection de certains dommages de l'ADN. Des protéines de cette voie de réparation sont en effet impliquées dans la signalisation cellulaire des dommages de l'ADN suivants (Li, 2008):

- Bases méthylées O₄-méthyle thymines (^{O₄-Met}T) et de O₆-méthyle guanines (^{O₆-Met}G). Ces bases sont produites par des agents méthylants comme le MNNG (pour Methylnitroinosoguanidine) ou le telozolomide. Des expériences par gel retard ont mis en évidence une forte affinité entre hMutS α purifié et un mésappariement ^{O₆-Met}G·T ou un appariement méthylé ^{O₆-Met}G·C et la structure cristallographique de MutS α complexé avec un mésappariement ^{O₆-Met}G·T a été résolue (Warren et al., 2007). Cette structure révèle que le mode de reconnaissance du mésappariement ^{O₆-Met}G·T est identique à celui d'un mésappariement G·T. Ces dommages peuvent être formés directement par un agent méthylant ou peuvent se former après réplication de

la base endommagée (Duckett et al., 1996). Les cellules humaines *msh6*^{-/-} et *mlh1*^{-/-} sont incapables de réparer ces lésions et deviennent résistantes aux agents méthylants;

- Incorporation de 6-thioguanine ou 5-fluorouracil (et leurs dérivés) comme substituts de nucléotides: le fluorouracile (5-FU) ou 5-fluorouracile fait partie des composés utilisés fréquemment en chimiothérapie notamment dans le traitement du cancer du côlon. En tant qu'analogue de la pyrimidine, le 5-FU est incorporé dans l'ADN et l'ARN, induisant au bout du compte l'arrêt du cycle cellulaire et l'apoptose. Ces analogues de bases peuvent former un mésappariement avec une thymine lors de la réplication, mésappariement qui est reconnu *in vitro* par hMutS α (Swann et al., 1996);

- Adduits formés par le cisplatine (ou *Cis*-diamminedichloroplatinum (II)): Le cisplatine est un composé utilisé en chimiothérapie dans le traitement de différents cancers tels que les sarcomes, carcinomes et lymphomes. Le cisplatine se fixe sélectivement sur les bases puriques de l'ADN (A ou G). les dommages induits sont pris en charge principalement par la voie de réparation NER (Nucleotide Excision Repair). Cette déformation inhibe la réplication et la transcription de l'ADN en ARN, et induit par ce biais la mort cellulaire. Le complexe hMutS α interagit *in vitro* avec l'adduit intrabrin 1,2-GpG (Duckett et al., 1996) et les cellules tumorales déficientes en MLH1 ou MLH2 montrent une résistance deux fois plus grande au cisplatine que celles complémentées par le chromosome portant le gène de la protéine manquante; il est à noter que la résistance au cisplatine en l'absence de MMR est moins forte que pour le 5-fluorouracile car les adduits créés par le cisplatine sont majoritairement pris en charge par le NER.

L'activité cytotoxique de ces composants fait qu'ils sont classiquement utilisés comme médicaments lors de chimiothérapie comme évoqué un peu plus haut; ainsi beaucoup de tumeurs présentant un déficit d'expression des protéines MMR deviennent dès lors, résistantes à ces traitements et sont alors difficiles à supprimer.

Deux modèles ont été proposés pour décrire le rôle du MMR dans la signalisation des dommages de l'ADN:

- Le modèle du cycle futile:

Selon ce modèle, le MMR aurait un seul et même rôle, celui de réparer les dommages de l'ADN. Dans ce modèle, suite au passage de la fourche de réplication, MutS α et MutL α reconnaissent la paire de base contenant le dommage présent sur le brin parental comme un mésappariement. Le complexe MutS α -MutL α active l'incision du brin néosynthétisé qui est dégradé puis resynthétisé, insérant au passage une base chimiquement intacte en face du nucléotide endommagé. Le dommage est donc toujours présent à l'issue de ce premier cycle et le complexe MutS α pourra alors de nouveau reconnaître le dommage et déclencher de façon répétée la réparation "futile" de la base présente sur le brin néosynthétisé. Il est proposé que ces cycles

répétés d'excision/resynthèse généreraient des brèches et des cassures dans l'ADN, trop nombreuses pour être réparées en totalité et déclencheraient au final l'apoptose (Li, 1999).

Pour tester ce modèle, des essais avec des cellules humaines exposées au 6-thioguanine ont été réalisés. Un arrêt en G2 du cycle cellulaire a été observé au niveau du second cycle cellulaire suivant l'exposition. L'arrêt du cycle cellulaire mesuré par cytométrie en flux est corrélé avec l'apparition de cassures simple-brin mesurées par l'essai "Comet". Ces résultats suggèrent que ce sont ces cassures qui activent le mécanisme d'apoptose (Yan et al., 2003). De son côté, l'équipe de P. Modrich a montré *in vitro*, que le MMR réparait un dommage ^{O6-Met}G·T porté par un plasmide si la base méthylée appartient au brin portant une incision, mais est incapable de le réparer si cette même base est positionnée sur le brin opposé (York and Modrich, 2006).

- Le modèle de signalisation directe

Dans ce modèle, le MMR aurait deux fonctions séparées: celle de réparation et celle de signalisation du dommage. Les complexes MutS α et MutL α reconnaîtraient alors le mésappariement induit chimiquement et activeraient directement le mécanisme de surveillance mettant en jeu les protéines p53-ATR-Chk1.

Différents travaux appuient ce second modèle. Une interaction physique entre la protéine MSH2 et la serine/thréonine protéine kinase ATR (Ataxia Telangiectasia and Rad3-related protein) a été observée par co-IP et cette interaction est essentielle pour la phosphorylation de SMC1 et Chk1, deux kinases intervenant dans la signalisation des dommages, suite à l'exposition de la cellule au MNNG (Wang and Qin, 2003). Plus récemment, il a été confirmé qu'ATR co-immunoprécipite avec MSH2 fixé à un mésappariement ^{O6-Met}G·T; la même étude a mis en évidence que MutL α est également nécessaire pour activer la phosphorylation de Chk1 (Yoshioka et al., 2006).

Les données biochimiques récentes semblent aller en faveur du modèle de signalisation directe mais il est difficile d'exclure l'un ou l'autre des modèles voire l'action de ces deux modèles.

Le MMR comme promoteur de stabilité génomique

Un rôle important des protéines du MMR dans les processus de recombinaison méiotique a été mis en évidence dès les années 1990 chez la levure et chez les mammifères (Santucci-Darmanin and Paquis-Flucklinger, 2003).

Chez la levure, il a été montré que la délétion des protéines homologues de MutS, Msh4 et Msh5 induit une diminution de la fréquence des enjambements (crossover en anglais) qui ont lieu lors de la ségrégation des chromosomes homologues au cours de la prophase I de la méiose. L'absence de crossover s'accompagne d'une

non-disjonction des chromosomes. Les protéines Msh4 et Msh5 forment un hétérodimère Msh4*Msh5. L'absence de ces protéines Msh4 et Msh5 ne s'accompagne pas de phénotype mutateur pour les souches de levure correspondantes ce qui suggèrent qu'elles ne participent pas à la voie de réparation MMR. Chez les mammifères, la construction des modèles animaux déficients pour les protéines homologues de MutS a permis de montrer le rôle majeur de certaines protéines Msh dans la recombinaison méiotique. Ainsi, les souris *msh2*^{-/-} et *msh6*^{-/-} sont fertiles; en revanche, des souris *msh5*^{-/-} (de Vries et al., 1999) et *msh4*^{-/-} (Kneitz et al., 2000) sont stériles, résultant d'un arrêt de la méiose à un stade précoce.

Dans la recombinaison méiotique, Mlh1 semble intervenir de préférence sous forme d'hétérodimère avec Mlh3 mais également avec Pms1 (ou PMS2 humain). Des mutants *mlh1Δ* ou *mlh3Δ* chez la levure présentent un défaut dans la formation des crossover (Hunter and Borts, 1997; Wang et al., 1999). Les souris mâles et femelles (Prolla et al., 1998) et (Lipkin et al., 2002) sont stériles mais le phénotype des souris *mlh1*^{-/-} est légèrement différent de celles *mlh3*^{-/-} car les souris déficientes en MLH1 présentent encore quelques nodules de recombinaison, nodules qui n'existent pas chez les souris déficientes en MLH3 (Kolas et al., 2005). Cela indique un rôle majeur de MLH3 dans la formation des crossover. Pms1/PMS2 joue également un rôle significatif puisqu'une délétion *pms1Δ* entraîne un défaut de viabilité des spores chez la levure. Les souris mâles *pms2*^{-/-} sont stériles tandis que les femelles *pms2*^{-/-} conservent leur fertilité. Enfin, les souris *exo1*^{-/-} sont stériles, que ce soit un mâle ou une femelle, indiquant un rôle essentiel d'Exo1 dans la méiose (Wei et al., 2003).

De nombreuses données de microscopie des cellules en méiose sont venues compléter ces dernières années les données sur le rôle des homologues MutS et MutL dans les différentes étapes de méiose (Santucci-Darmanin and Paquis-Flucklinger, 2003). Il ressort de ces études que le complexe MSH4-MSH5 s'associe aux chromosomes appariés dès les étapes précoces de la prophase I à un stade appelé zygotène et reste associé jusqu'au stade plus tardif de la prophase appelé pachytène. Ces stades encadrent les mécanismes de recombinaison méiotique qui ont lieu entre l'initiation des cassures double brin et la formation des crossover. On pense aujourd'hui que ce complexe permet de stabiliser les jonctions de Holliday qui se forment lors de la recombinaison (Snowden et al., 2004). L'ordre précis du recrutement et le rôle des complexes MLH1*MLH3 et/ou MLH1*PMS2 ainsi que d'EXO1 n'est pas encore bien compris. Cependant, il est à noter que Mlh3 contient dans sa séquence les motifs d'endonuclease observés sur les protéines Pms1 de levure et de PMS2. Des mutations dans l'un des motifs endonuclease de Mlh3 provoquent un défaut de recombinaison méiotique chez la levure, équivalent à une délétion *mlh3Δ* (Nishant et al., 2008). Ces données suggèrent une participation de Mlh3 dans la recombinaison méiotique notamment à travers son activité endonuclease.

Les protéines du MMR ont également été mises en évidence comme influençant l'efficacité de la recombinaison mitotique. Des délétions *msh2Δ* et *msh6Δ* chez la levure présente une augmentation de recombinaison entre des séquences divergentes (Goldfarb and Alani, 2005). Ainsi, le MMR semble essentiellement prévenir l'appariement de séquences non-homologues.

Le MMR comme promoteur de variation génétique

De façon paradoxale, les protéines du MMR sont également impliquées dans les mécanismes permettant d'accroître la diversité de certains gènes, en particulier des anticorps. Au sein des cellules B, les immunoglobulines subissent des mutations et des réarrangements de gènes. Le premier mécanisme est important pour la génération d'anticorps très affins pour leur antigène. Le second mécanisme joue un rôle dans la fonctionnalisation des immunoglobulines.

Le premier de ces mécanismes est l'hypermutation somatique (SHM pour Somatic Hypermutation en anglais). Le SHM permet d'atteindre un taux de mutations de 10^{-3} à 10^{-5} sur les gènes d'immunoglobulines (Rajewsky et al., 1987). Les mutations correspondent essentiellement à des substitutions d'une paire de base dans les régions variables des immunoglobulines. Le rôle des protéines du MMR dans le SHM a été proposé notamment dans le cadre de l'étude de souris "KO". Des souris *msh2^{-/-}* et *msh6^{-/-}* présentent une diminution d'un facteur 5 et par 2, respectivement, du taux de mutations au niveau des régions variables des immunoglobulines par rapport à des souris sauvages (Martomo et al., 2004; Rada et al., 1998). Les souris "KO" *mlh1^{-/-}* et *pms2^{-/-}* présentent également une variation du taux de mutations, *i.e.* une diminution d'un facteur 2 et 1.5, respectivement, par rapport aux souris sauvages (Phung et al., 1999; Winter et al., 1998). Plus récemment, des souris *mlh3^{-/-}* ont été étudiées. Elles présentent également une élévation du taux de mutations des régions variables des immunoglobulines (Wu et al., 2006).

Il a été montré que dans les étapes initiales de la voie SHM, l'enzyme AID (pour Activation-Induced cytidine Deaminase) retire le groupement amine d'une cytosine changeant cette base en uridine. L'AID crée ainsi un mésappariement G·U. à cette étape, deux scénarios sont observés (Li et al., 2004). Soit le mésappariement est transformé en site abasique par la protéine UNG (pour "uracil N-glycosylase") puis pris en charge par le BER (pour Base Excision Repair). Dans ce cas là, il n'y a pas de mutation introduite. Soit, le mésappariement est reconnu par MutSα et on observe alors l'introduction d'une mutation par un mécanisme non encore élucidé. L'équipe de J. Jiricny a proposé un modèle à partir d'expériences *in vitro* de réparation MMR de plasmides contenant un ou deux mésappariements de type U·G ou G·T (Schanz et al., 2009): comme il a été dit précédemment, le MMR n'est pas capable de prendre en charge un mésappariement porté par un plasmide clos puisque le MMR nécessite une incision comme site de discrimination. En revanche, si le plasmide contient un second mésappariement U·G situé à 54bp d'un premier mésappariement G·T, ce dernier

mésappariement est corrigé en A·T par le MMR. Il a été montré que la réparation n'était plus possible en présence d'inhibiteurs de glycosylases. Il est donc supposé que le mésappariement U·G est pris en charge par le BER et l'action de ce mécanisme introduit une incision. Cette incision permet à MutS α *MutL α de recruter EXO1 afin de réparer le mésappariement G·T. Dans ce cas, la resynthèse effectuée par un ADN polymérase fidèles.

Cependant, le MMR mis en œuvre dans la voie SHM dépend non pas des ADN polymérase fidèles comme c'est le cas du MMR post-répliatif, mais de l'ADN polymérase translesionnelle Pol η , qui est connue pour introduire une base non-complémentaire en face d'un T. L'action de cette polymérase va induire alors une mutation au niveau d'un paire de base A·T. Il a été montré que la polymérase Pol η humaine interagit avec MutS α humain (interaction démontrée par pull-down et co-IP (Wilson et al., 2005)) et avec MLH1 humain (interaction démontrée par pull-down et co-IP (Kanao et al., 2009)).

Les protéines du MMR sont également impliquées dans un autre mécanisme de diversification des immunoglobulines, à savoir la commutation isotypique (CSR pour Class Switch Recombination en anglais). Ce mécanisme intervient sur les régions constantes de la chaîne lourde des immunoglobulines. Le changement de régions constantes des immunoglobulines confère à ces dernières vers des fonctions effectrices différentes tout en gardant la spécificité du site de reconnaissance antigénique. La commutation isotypique permet, par recombinaison de cassures double brin, de remplacer la région constante C μ initialement présente par d'autres séquences de régions constantes (C γ , C ϵ , C δ ou C α). L'étude des souris *msh2*^{-/-} a montré que ces souris *msh2*^{-/-} et *msh6*^{-/-} présentent 2 à 10 fois moins de recombinaisons lors de la commutation isotypique (Ehrenstein and Neuberger, 1999; Schrader et al., 1999). Les souris *mlh1*^{-/-} (Schrader et al., 1999) ou *pms2*^{-/-} (Ehrenstein et al., 2001) présentent également une réduction d'un facteur allant de 2 à 4 du taux de recombinaison.

La commutation isotypique présente des étapes communes avec la voie SHM. Cette voie met en jeu, les enzymes AID et UNG ainsi que les endonucléases du BER, aboutissant à la formation d'une cassure simple-brin. Le mécanisme permettant de passer de cette incision d'un brin à une cassure double-brin nécessaire à la recombinaison, reste aujourd'hui partiellement compris. Des expériences de LM-PCR (pour ligation-mediated polymerase chain reaction) montrent que des cellules B déficientes en MSH2 ou MLH1 présentent un nombre restreint de cassures double-brin (Schrader et al., 2007). De plus, il a été montré que des cellules B *msh2*^{-/-} n'effectuent de recombinaisons qu'entre incisions proches alors que dans les cellules sauvages, la recombinaison intervient entre nucléotides plus éloignés (Ehrenstein and Neuberger, 1999). Ces résultats suggèrent que des cytosines proches de moins de 5bp forment spontanément des cassures double-brins (Pierce et al., 2001). Il a été proposé que si les incisions faites par le BER sont trop éloignées pour déstabiliser l'ADN

double brins, le mésappariement G·U est reconnu par MutS α et MutL α qui vont alors recruter EXO1. L'exonucléase EXO1 peut dégrader le brin entre l'incision précédente et le mésappariement et le brin peut alors être synthétisé entre les deux incisions à travers Pol η . La polymérase translésionnelle Pol η est nécessaire dans le CSR pour passer la barrière du site abasique non pris en charge par les polymérases fidèles. Ceci aboutit à une cassure double-brin permettant la recombinaison en amont des régions constantes puis la réparation de la cassure par le mécanisme NHEJ (pour NonHomologous End Joining).

vi. Impact d'une déficience du MMR et syndrome de Lynch

Le syndrome de Lynch et ses conséquences cliniques

Ce syndrome, également appelé syndrome HNPCC (Hereditary NonPolyposis Colorectal Cancer) a été décrit pour la première fois en 1913 par A.S. Warthin, en observant les prédispositions de certaines familles à développer des cancers colorectaux (Warthin, 1913). Cette étude a été complétée en 1966 par H.T. Lynch, qui étudia deux familles sur cinq générations. La première famille présente 7 cas de cancer colorectal, 2 cas de cancer de l'estomac, 7 cas de cancer de l'utérus et 12 autres cas de cancer. La seconde famille présente 5 cas de cancer colorectal, 1 cas de cancer de l'estomac, 5 cas de cancer de l'utérus et 5 autres cas de cancer. Ces données ont suggéré l'existence d'une transmission autosomique dominante d'une prédisposition familiale au cancer (Lynch et al., 1966). Une transmission autosomique dominante est une maladie génétique où l'allèle contenant le gène muté est sur un chromosome non sexuel et où la présence d'un seul allèle muté est suffisante pour que la maladie s'exprime.

Au début des années 90, cette prédisposition a été identifiée comme liée à un mécanisme MMR déficient; les tumeurs présentent pour la grande majorité, une instabilité des séquences répétées en nucléotides appelées microsatellites (ou MSI pour Microsatellite Instability). Les MSI concernent notamment des protéines qui présentent au sein de leur région codante des répétitions de nucléotide. Il s'agit par exemple du récepteur de classe II du TGF β (TGF β -RII), de la protéine BAX impliquée dans l'apoptose, ... (cf. § *i*). Cette instabilité a pour conséquence une augmentation du taux de mutation et une accumulation d'anomalies dans les gènes concernés. Il a été proposé que la prépondérance de cancer au niveau du colon aurait pour origine une mutation particulièrement importante des gènes impliqués dans la croissance de l'épithélium du côlon (Poulogiannis et al., 2010). Ainsi, le syndrome de Lynch est la forme la plus fréquente de prédisposition héréditaire aux cancers colorectaux. Ces cancers représentent 2 à 7% de l'ensemble des cancers colorectaux, soit environ 1000 nouveaux cas par an en France. Une des caractéristiques de ces cancers est qu'ils se déclarent à un âge peu avancé (42 ans en moyenne contre 65 ans pour des cancers colorectaux sporadiques (Jacob and Praz, 2002)). Ces cancers représentent 14% des cancers colorectaux détectés avant 50 ans (Lynch and de la

Chapelle, 2003). Bien que prépondérants, ces cancers n'atteignent pas seulement le côlon mais également l'endomètre et dans une moindre mesure, les ovaires, l'estomac, l'intestin grêle ainsi que le cerveau et la peau.

Il existe plusieurs études concernant les risques de développer un cancer lorsqu'on est atteint du syndrome de Lynch. Une des dernières études en date est l'étude appelée ERISCAM (Estimation des RISques de CANcer chez les porteurs de mutation des gènes MMR). Cette étude regroupe les données de 537 familles françaises avec une mutation d'un gène MMR identifiée, en utilisant une méthode statistique corrigeant le biais de sélection des familles. Cette étude montre que le risque cumulé de développer un cancer à 70 ans est de 45% pour un homme de 20 ans et 54% pour une femme de 20 ans. Ce risque dépend du gène portant la mutation. Le risque spécifique est d'environ 35% pour un cancer colorectal et de 28% pour un cancer de l'endomètre. Le risque de développer un autre cancer associé au syndrome de Lynch est de l'ordre de 10-15%.

Comme H.T. Lynch l'avait proposé, le syndrome HNPCC est un syndrome héréditaire dominant, c'est-à-dire que chaque cellule somatique possède un allèle muté et un allèle sauvage. Le développement d'une tumeur fait suite à l'inactivation somatique de l'allèle sauvage par délétion, mutation ponctuelle ou hyperméthylation du promoteur (Potocnik et al., 2001).

Deux cas particuliers du syndrome de Lynch existent de façon plus exceptionnelle: d'une part, il existe un certain nombre de cancers sporadiques liés aux gènes du MMR. Dans 90% des cas, ils ne présentent pas de mutations dans les gènes du MMR et sont liés à une hyperméthylation du promoteur de MLH1, qui a pour conséquence la non-expression de la protéine (Kuismanen et al., 2000). D'autre part, dans de rares cas, les deux allèles d'un même gène du MMR portent une mutation. Le syndrome de Turcot est le cas le plus connu d'homozygotie de ce type et est caractérisé par l'apparition à la fois d'une tumeur au cerveau et d'un cancer colorectal qui apparaissent chacun en moyenne avant 15 ans. Ces mutations homozygotes n'ont été décrites à ce jour que pour le gène codant PMS2, avec des parents qui ne développent pas la plupart du temps de cancer HNPCC (De Rosa et al., 2000).

Le syndrome de Lynch associé à une déficience du MMR

Bien que les conséquences cliniques soient connues depuis quelques dizaines d'années, les causes génétiques n'ont été élucidées que récemment. Aujourd'hui, 2784 mutations sur les gènes du MMR ont été associées au syndrome de Lynch d'après la base de données établie par l'InSiGHT (International Society for Gastrointestinal Hereditary Tumours - <http://www.insight-group.org/>).

Les 1^{ères} mutations ont été identifiées sur les gènes de *mlh1* (Lindblom et al., 1993) et *msh2* (Peltomäki et al., 1993). Les mutations sur ces deux gènes représentent 73% des mutations identifiées (39% sur *mlh1* et 34% sur *msh2*).

Les mutations sur le gène *msh2* ont un effet très marqué puisqu'un homme porteur d'une de ces mutations a 63% de risques cumulés de développer un cancer colorectal avant 70 ans tandis qu'une femme a un risque de 68% (Jang and Chung, 2010). L'âge moyen de l'apparition de ces tumeurs est de 44 ans. Parmi toutes les mutations, environ 80% aboutissent à une protéine MSH2 tronquée. Ces protéines tronquées viennent de petites insertions ou délétions décalant le cadre de lecture (~50% des cas), d'une substitution codant pour un codon stop (~25% des cas), de larges délétions dans le gène (~25% des cas) (Woods et al., 2007). L'ensemble des autres mutations concernant *msh2* sont des mutations faux-sens ayant comme répercussion un changement d'un acide aminé. L'impact sur la structure ou la fonction d'une grande partie de ces variants ponctuels n'a pas encore été caractérisé à ce jour.

Les mutations sur le gène *mlh1* ont également un effet très important puisque les risques cumulés d'avoir un cancer colorectal à 70 ans sont de 65% pour un homme et de 53% pour une femme, avec un âge moyen d'apparition des tumeurs de 43 ans (Jang and Chung, 2010). La différence entre hommes et femmes n'est pas comprise à ce jour. Près d'un tiers des mutations identifiées sont des mutations ponctuelles qui affectent les fonctions de MLH1 et comme pour la protéine MSH2, rares sont les mutations caractérisées d'un point de vue fonctionnel. Les autres mutations correspondent, comme précédemment, à l'expression d'une protéine tronquée suite à l'apparition d'un codon stop, d'une insertion/délétion, décalant le cadre de lecture, ou d'une délétion d'une grande séquence du gène.

D'autres gènes, ayant un impact un peu moins fort, ont été identifiés par la suite. (Miyaki et al., 1997) comptent 539 mutations à ce jour, soit 19% de l'ensemble des mutations ponctuelles associées à des cas de cancer HNPCC. *pms2* (Hendriks et al., 2006) comptent 184 mutations répertoriées, représentant alors 7% des mutations.

Il a été proposé que cette faible occurrence des mutations du gène *msh6* reflète la redondance partielle dans les rôles joués par les protéines MSH6 et MSH3. De fait, l'apparition des tumeurs est plus tardive (autour de 55-60 ans). De plus, le phénotype associé à ces mutations est assez particulier, puisque le risque cumulé d'avoir un cancer colorectal avant 70 ans est de 69% pour un homme mais seulement de 30% pour une femme. En revanche une femme possédant une de ces mutations a 71% de chance d'avoir un cancer de l'endomètre (presque deux fois plus que pour une mutation sur *msh2*). Malgré la recherche de phénotype associé, aucune mutation sur *msh3* n'a été associée à des MSI (Huang et al., 2001).

Les mutations portant sur *pms2* ont longtemps été sous-estimées car elles prennent souvent part à des phénotypes intermédiaires (hormis dans les cas

exceptionnels d'homozygotie comme évoqué précédemment) et sont réparties au milieu de gènes incomplets appelés pseudogènes; le risque cumulé à 70 ans est de l'ordre de 15-20% pour les cancers colorectaux ou de l'endomètre (Jang and Chung, 2010). Il a été proposé qu'en présence de mutations de PMS2, la protéine MLH3 pourrait compenser en partie le rôle de la protéine PMS2 (Chen et al., 2005). Il est intéressant de noter que plus de la moitié des mutations prédites code pour une mutation ponctuelle sur la protéine.

Le rôle de *mlh3* dans le développement de ce type cancer est encore aujourd'hui assez ambigu. Certaines mutations ont été identifiées (35 au total) mais aucune n'a clairement été associée au développement de tumeurs présentant des MSI. Des analyses ont également été effectuées sur *pms1* (Liu et al., 2001) et *exo1* (Alam et al., 2003), aucune mutation sur ces deux protéines n'a été associée à des instabilités MSI.

1.1.2 MUTL ET SES HOMOLOGUES

i. Les homologues de MutL chez les eucaryotes et leurs fonctions dans le MMR

Comme évoqué précédemment, il existe 4 homologues de MutL chez les eucaryotes: Mlh1 et Mlh3 (MutL homologue 1 et 3) ainsi que Pms2 (nommée PMS2 chez l'homme et Pms1 chez la levure) et Pms1 (nommée PMS1 chez l'homme et Mlh2 chez la levure) pour post-meiotic segregation 1 et 2. Ces quatre protéines forment trois hétérodimères MutL α , MutL β et MutL γ ayant comme unité commune Mlh1 (Wang et al., 1999).

Le complexe MutL α est l'hétérodimère principal de la voie du MMR post-répliatif. En effet, des souris *mlh1*^{-/-} et *pms2*^{-/-} montrent une forte susceptibilité à développer des tumeurs: 50% des souris meurent au cours des 6 premiers mois si elles sont déficientes en MLH1 et au cours des 9-10 premiers mois si elles sont déficientes en PMS2 (Prolla et al., 1998). Il est intéressant de noter que la protéine PMS2 semble avoir besoin de la protéine MLH1 pour se stabiliser car les souris *mlh1*^{-/-} n'expriment pas la protéine PMS2 (Yao et al., 1999). Au niveau moléculaire, ces phénotypes observés ont d'abord été expliqués par un rôle de coordinateur, similaire à celui de MutL chez les bactéries (Jiricny and Nystrom-Lahti, 2000). Mais il est clair aujourd'hui que ce complexe a une autre fonction tout aussi importante puisque, comme il a été mentionné auparavant, MutL α possède une activité endonucléase aussi bien chez la levure que chez l'humain (Kadyrov et al., 2006; Kadyrov et al., 2007). L'activité endonucléase de MutL α explique le rôle essentiel que joue ce complexe dans la voie MMR lorsque la cassure simple brin se trouve en 5' du mésappariement. Dans ce cas de figure, Exo1 avec sa polarité 5'-3' ne peut exciser le brin néosynthétisé. L'incision par MutL α en 3' permet de charger Exo1 de l'autre

côté du mésappariement puis à Exo1 de dégrader le brin néosynthétisé (**Figure 1 - 1**). Cette activité de MutL α s'exerce aussi bien en présence de mésappariements base/base reconnus par MutS α (Constantin et al., 2005) qu'en présence d'IDLs reconnus par MutS α ou MutS β (Zhang et al., 2005).

Les souris mâles et femelles *mlh1*^{-/-} sont stériles et seules les souris mâles *pms2*^{-/-} sont stériles (Baker et al., 1995). Ces résultats indiquent que MutL α intervient dans la recombinaison méiotique. Il est proposé aujourd'hui que MutL α interviendrait lors de la reconnaissance de séquences non parfaitement homologues, appelées homéologues par Msh4*Msh5 plutôt que lors des événements favorisant les échanges réciproques entre chromosomes. Ceci est confirmé par le faible nombre de mutations impliquées dans le syndrome de Lynch et localisées sur le gène *pms2*.

Le complexe majoritairement impliqué dans la recombinaison méiotique est le complexe MutL γ composé de Mlh1 et Mlh3. Les souris *mlh3*^{-/-} présentent des anomalies dans la ségrégation méiotique des chromosomes, et une apoptose importante des cellules germinales. Ces souris sont donc stériles comme les souris *mlh1*^{-/-}. Il a été observé que la protéine Mlh3 est stable même en l'absence de Mlh1 et qu'elle est localisée sur les foyers de recombinaison à une étape antérieure à celle de fixation de Mlh1 (Lipkin et al., 2002). Au niveau biochimique, MLH3 a été montré comme interagissant avec MSH4 par co-IP avec des extraits cellulaires de souris (Santucci-Darmanin et al., 2002). Le rôle exact de MutL γ dans la recombinaison méiotique n'est pas bien établi. L'activité endonucléase de Mlh3 au sein du complexe MutL γ est nécessaire pour la recombinaison méiotique. Il a été proposé que le complexe Mlh1*Mlh3 serait impliqué dans la formation des intermédiaires de recombinaison intervenant entre la reconnaissance des jonctions de Holliday par Msh4*Msh5 et la réparation des cassures double-brin (Nishant et al., 2008).

Le rôle du complexe MutL γ dans le MMR mitotique est mal défini. Les souris *mlh3*^{-/-} présentent une susceptibilité à développer des lymphomes, qui se déclarent en général à un âge plus tardif que pour des souris *mlh1*^{-/-} ou même *pms2*^{-/-} (Chen et al., 2005). Ces souris présentent une instabilité au niveau des MSI, 9 fois plus élevée que pour une souris sauvage (mais 2 fois moins élevée que les souris *pms2*^{-/-}). Les données biochimiques ont montré que MLH3 interagit lors d'essais par pull-down avec MSH3 (Charbonneau et al., 2009). Ce résultat suggère une fonction MSH3-dépendante dans la réparation des grandes boucles d'insertion/délétion. Le rôle de MLH3 dans le système MMR post-réplcatif a en effet été confirmé par des essais de réparation *in vitro* sur des extraits cellulaires humains. De façon intrigante, la surexpression de MutL γ augmente de 20% l'efficacité de réparation d'un mésappariement base/base G·T ou l'insertion d'un nucléotide. En revanche aucune augmentation du taux de réparation n'est observée pour une insertion de deux nucléotides (Cannavo et al., 2005).

A ce jour, il est démontré que le niveau d'expression de MLH3 dans les cellules somatiques est très faible, à savoir environ 60% inférieur à celui de PMS2, que ce soit au niveau de l'expression de la protéine (Cannavo et al., 2005) que du taux d'ARN messagers correspondant (Charbonneau et al., 2009). Ce taux est également bien inférieur à celui existant dans les cellules germinales où Mlh3 est impliqué lors de la méiose. On en conclut qu'à ce taux d'expression, la réparation des mésappariements par l'intermédiaire de MutL γ , déjà peu efficace par surexpression, est vraisemblablement très faible en conditions physiologiques. S'il existe, le rôle de MutL γ après le passage de la fourche de réplication, serait probablement celui de voie de sauvegarde (ou back-up en anglais) par rapport au complexe MutL α .

Mlh1 forme un dernier hétérodimère, MutL β , avec PMS1 chez l'homme (Raschle et al., 1999) et Mlh2 chez la levure (Wang et al., 1999). La formation des hétérodimères de levure et humains a été démontrée par double-hybride et co-IP. Très peu de données sont disponibles quant à la fonction de cet hétérodimère. Chez *S. cerevisiae*, des souches *mlh2 Δ* possèdent une sensibilité augmentée d'un facteur 2-3 au cisplatine et au carboplatine (Durant et al., 1999). Ces données suggèrent un rôle de PMS1/Mlh2 dans la signalisation des dommages de l'ADN. Cependant ce rôle n'a pas été confirmé depuis.

Le rôle de MutL β dans le MMR post-réplicatif n'est pas démontré. Les souris *pms1*^{-/-} sont fertiles et se développent normalement, sans susceptibilité au cancer. Au niveau moléculaire, ces souris ont une déficience dans la réparation des répétitions d'un nucléotide mais ne présentent pas d'instabilité au niveau de microsatellites dinucléotidiques. D'autre part, le complexe MLH1-PMS1 n'a pas, *in vitro*, d'activité de réparation des mésappariements base/base ou d'IDLs de 1 ou 2 nucléotides (Raschle et al., 1999). Cependant une étude génétique chez la levure utilisant la réversion d'un allèle *lys2* muté au sein d'une répétition de 10 nucléotides a montré que des souches *mlh2 Δ* présentent une légère augmentation du taux de réversion (Harfe et al., 2000). Ceci met en évidence un rôle de MutL β dans le MMR post-réplicatif, rôle qui, bien que non-essentiel, semble être complémentaire à celui de Mlh3 dans un système MMR Msh3-dépendant.

Abondant dans le sens d'une complémentarité entre Mlh2/PMS1 et Mlh3, une seconde étude a montré que des levures *mlh2 Δ* ne présentent pas de diminution de formation des crossover; en revanche, des cellules *mlh2 Δ mlh3 Δ* présentent une augmentation significative du taux de crossover par rapport à des cellules *mlh3 Δ* (Abdullah et al., 2004). Les auteurs ont émis l'hypothèse que Mlh2/PMS1 pouvait servir d'aiguilleur vers un mécanisme Msh4-dépendant ou vers un système de Mus81-dépendant en absence de Msh4 (Mus81 est une endonucléase, structure spécifique, des jonctions formées lors des crossover).

En conclusion, les trois hétérodimères MutL α , MutL β et MutL γ qui possèdent Mlh1 comme sous-unité commune assurent des fonctions différentes qui se recouvrent partiellement et permettent de palier provisoirement un dysfonctionnement d'un des complexes. Les fonctions principales assurées par ces hétérodimères sont la voie de réparation MMR réplicatif avec principalement MutL α et la voie de recombinaison méiotique avec principalement MutL γ .

ii. Organisation des domaines chez les homologues de MutL et leurs rôles dans son fonctionnement

Aucune structure 3D de protéine MutL bactérienne entière ou d'homologue eucaryote de MutL entier n'a encore été résolue. Mais les structures cristallographiques de domaines de MutL et de domaines d'homologues de MutL ont été résolues. Ces données structurales et de nombreuses données biochimiques, ont montré que l'ensemble de cette famille MutL possède une architecture globale identique avec 3 parties: (i) un domaine N-terminal (d'environ 300 résidus) conservé, portant une activité ATPase et impliqué dans l'interaction avec l'ADN ainsi qu'avec MutS ou ses homologues; (ii) un domaine C-terminal moins conservé en séquences et comprenant notamment le domaine d'homodimérisation des protéines MutL bactériennes ou d'hétérodimérisation des homologues eucaryotes de MutL; (iii) une région charnière, prédite comme désordonnée et reliant les deux domaines précédents. La longueur de cette région varie de quelques acides aminés chez des organismes comme *A. aeolicus* et *D. radiodurans* à environ 150 acides aminés chez la levure et l'homme.

Le domaine N-terminal de MutL bactériens et des homologues eucaryotes de MutL

Le domaine N-terminal de MutL (ou NTD pour N-terminal domain) et de ses homologues présente un repliement identique à ceux des protéines de la famille ATPases GHKL (pour Gyrase, Hsp90, Histidine kinase et MutL). Cette famille comprend également la topoisomérase de type II (Dutta and Inouye, 2000). Cette famille se caractérise par 4 motifs très conservés, impliqués dans l'activité ATPase commune à ces protéines. Ces 4 motifs sont présents dans les domaines N-terminaux des protéines MutL, Mlh1, Pms1/PMS2, Mlh3 et Pms1.

Plusieurs structures cristallographiques de la région N-terminale de MutL d'*E. coli* ont permis de mieux comprendre le fonctionnement et le rôle de cette activité ATPase. Ces structures 3D, des expériences de chromatographie par exclusion de taille et de pontage ont montré que le domaine N-terminal de la protéine en présence d'ADP ou en l'absence de nucléotide est un monomère (**Figure 1 -8(A)**) (Ban and Yang, 1998a). La fixation d'un analogue de l'ATP non-hydrolysable comme l'ADPnP provoque une transition de la forme monomérique vers une forme dimérique (**Figure 1 -8(B)**) (Ban et al., 1999). La fixation d'un nucléotide induit un changement conformationnel, notamment au niveau de 3 boucles:

- la boucle L1 constituée des 19 premiers résidus de l'extrémité N-terminale;
- la boucle L2 constituée des résidus 150 à 162
- la boucle L3 constituée des résidus 299 à 313.

Ces boucles comportent des acides aminés très conservés mais n'ont pu être modélisées dans la densité électronique de la protéine cristallisée sans nucléotide, de par leur mobilité. En revanche, ces trois boucles acquièrent une structure bien définie dans la structure cristallographique du domaine N-terminal complexé avec l'ADPnP. Les boucles L1 et L3 sont directement impliquées dans l'interaction avec le nucléotide, notamment à travers la lysine K307 qui interagit spécifiquement avec un groupement phosphate de l'ADPnP. L'autre partie du site de fixation est composé des motifs I, II et IV caractéristiques de la famille GHKL et forment des liaisons hydrogène avec l'adénosine. Enfin, un "couvercle" constitué du motif III des ATPases GHKL (résidus 74 à 100), vient compléter ce site actif.

La structure cristallographique de la région N-terminale en présence d'analogue d'ATP met en évidence une dimérisation de cette région. Cette dimérisation est rendue possible par la structuration des boucles L1, L2 et L3. L'interface hydrophobe de la surface de dimérisation est constituée d'un feuillet β à l'extrémité de la boucle L1 et de la région équivalente du second monomère (**Figure 1 - 8(B)**). Cette interface est constituée également du reste de la boucle L1 qui adopte une structure en hélice et interagit avec l'hélice C du couvercle du second monomère. La dimérisation est renforcée également par la structuration de la boucle L2.

Il a été montré *in vitro* qu'en présence d'ATP et de MutS, le domaine N-terminal était suffisant pour activer l'endonucléase MutH et que cette activation est liée aux changements conformationnels observés lors de la fixation de l'ATP (Ban and Yang, 1998a) (**Figure 1 - 8(C)**). En effet, une des premières fonctions identifiées pour MutL est sa fonction de coordinateur entre les activités de reconnaissance de MutS et d'incision de MutH. Par combinaison de méthodes bioinformatiques et de pontage, il a été montré que la région N-terminale de MutL est importante pour l'interaction avec MutS (Plotz et al., 2006). La région d'interaction est proche du site ATPase. Une étude par résonance plasmonique de surface a montré que la fixation de l'ATP sur MutL n'est pas requise pour l'interaction avec MutS (Acharya et al., 2003). En effet, le site d'interaction se trouve à la surface de MutL, dans une région accessible aussi bien lorsque la protéine MutL est sous sa forme monomérique que dimérique. *A contrario*, l'activation de l'endonucléase MutH par MutL est ATP-dépendante. L'activité endonucléase de MutH seule est relativement faible mais l'interaction avec le dimère du domaine N-terminal de MutL augmente l'efficacité d'incision de l'endonucléase (Acharya et al., 2003). L'activation spécifique de MutH par la forme dimérique de MutL a été expliquée par la cartographie du site d'interaction de MutH sur MutL. Les quatre résidus de MutL (N169, A251, Q314 et L327) identifiés comme

essentiels pour l'interaction avec MutH suggèrent que MutH se fixe entre les deux sous-unités du dimère MutL. (Giron-Monzon et al., 2004) (**Figure 1 - 8(C)**).

La formation du dimère de la région N-terminale de MutL crée une cavité chargée positivement au niveau du second domaine α/β . Ce site est important dans la fixation de l'ADN par MutL, de préférence sous forme simple brin (comme après le passage d'une exonucléase) (Robertson et al., 2006). En accord avec cette interaction, le mutant R266E de la cavité de MutL perd la capacité à lier l'ADN simple brin et présente un fort phénotype mutateur dans des essais de complémentation sur plasmide (Robertson et al., 2006). D'autres études suggèrent que la fixation de l'ADN simple brin proposée par cette dernière étude n'est pas pertinente d'un point de vue physiologique (Drotschmann et al., 1998) (Acharya et al., 2003). Une étude récente par molécule unique observe également l'interaction avec un ADN simple brin mais uniquement à basse force ionique (Park et al., 2010). Il semble de plus en plus clair que l'interaction des protéines MutL avec l'ADN fait intervenir plusieurs sites aussi bien chez la bactérie (Park et al., 2010) que chez les eucaryotes (Hall et al., 2001). Un second site pourrait faire intervenir le domaine C-terminal de MutL. En accord avec cette dernière proposition, des expériences de microscopie à force atomique sur le complexe MutL α entier montrent que la fixation de l'ATP sur le domaine N-terminal provoque la condensation de l'homodimère MutL entier suite au repliement sur lui-même de la région déstructurée entre les domaines N- et C-terminaux (Sacho et al., 2008). Il a été proposé que cette condensation permettrait à l'homodimère MutL entier d'entourer le brin d'ADN entre ses deux domaines N- et C-terminaux.

La fixation de MutL sur l'ADN simple brin et les changements conformationnels qui lui sont associés pourraient permettre de réguler l'activité des protéines agissant en aval dans la voie MMR comme l'hélicase UvrD (Guarne et al., 2004).

Chez les eucaryotes, la conformation et les fonctions de la région N-terminale des homologues de MutL sont similaires à celles observées pour MutL d'*E. coli*. Plusieurs structures cristallographiques de domaines NTD de MutL eucaryotes ont été résolues. Trois structures cristallographiques sont déposées dans les banques de données des structures : PMS2 seul (Guarne et al., 2001), Pms1 seul (Arana et al., 2010) et MLH1 seul (*non publié*). En revanche, aucune structure d'hétérodimère de la région N-terminale de MutL eucaryote n'est connue. Ces structures se superposent bien avec la région N-terminale de MutL d'*E. Coli* (rmsd de 2.2Å sur 261 atomes C $_{\alpha}$ pour PMS2, de 1.8Å sur 234 atomes C $_{\alpha}$ pour Pms1 et de 1.8Å sur 279 atomes C $_{\alpha}$ pour MLH1).

Le domaine N-terminal de PMS2 est monomérique dans la structure cristallographique sans adénosine mais les boucles L2 et L3 sont déjà partiellement structurées (**Figure 1 - 8(D)**). De plus, l'ajout d'ATP γ S sur PMS2 ou d'AMPPNP sur Pms1 ne provoque pas la dimérisation contrairement à ce qui est observé pour

ecMutL. En revanche, la structure concernant le domaine de MLH1 lié à l'ATP est un homodimère dont l'agencement est similaire à celui de MutL, notamment au niveau des boucles qui forment l'interface (**Figure 1 - 8(E)**). Le partenaire principal de la région N-terminal de Mlh1 est la région équivalente de Pms1. L'équipe de R. Liskay a observé par double-hybride l'interaction entre les régions N-terminale de Mlh1 et Pms1 de *S. cerevisiae*. Cependant, il a également été montré par ultracentrifugation analytique que la protéine Mlh1 de levure seule forme des homodimères avec une constante de dissociation de 3 μ M contre 83nM pour le complexe Mlh1*Pms1 (Shcherbakova et al., 2001). La structure cristallographique de la région N-terminale de Mlh1 suggère que cette région contribue en partie à l'homodimérisation observée de Mlh1 entière. De façon identique au mécanisme procaryotique, la fixation de l'ATP induit des changements conformationnels permettant la dimérisation du domaine N-terminal de Mlh1 avec le domaine N-terminal de Pms1 (Tomer et al., 2002; Tran and Liskay, 2000).

Cependant, plusieurs observations chez la levure (Hall et al., 2002) et l'humain (Tomer et al., 2002) prouvent que l'affinité de Mlh1 pour l'ATP est 10 fois supérieure à celle de Pms1/PMS2; cette asymétrie a également été observée directement par microscopie à force atomique sur le complexe MutL α entier (Sacho et al., 2008). En l'absence d'adénine, MutL α existe essentiellement en conformation étendue, *i.e.* que le bras reliant les domaines N- et C-terminaux est étendu. L'ajout d'ATP à faible concentration induit une augmentation des conformations semi-condensées où un des deux bras seulement se replie. Enfin à forte concentration d'ATP, la forme condensée devient prédominante. Ceci démontre un fonctionnement séquentiel de l'hétérodimère MutL α où différents états co-existent, permettant probablement de réguler la fixation du complexe sur l'ADN ainsi que l'interaction avec d'autres protéines effectrices.

La région N-terminale de MutL α est également impliquée dans l'interaction avec MutS α mais la fonction de cette interaction est différente puisqu'aucun homologue de MutH n'est connu chez les eucaryotes. Au sein de l'hétérodimère MLH1*PMS2 humain, le domaine N-terminal de MLH1 a été défini comme étant suffisant pour l'interaction avec le complexe MutS α par des expériences de co-IP (Plotz et al., 2003). Quatre positions de MLH1 essentielles pour l'interaction MLH1-MutS α ont été identifiées (H112, R127, A128 et Y130). La localisation de ces positions indique que MutL et MLH1 utilise les mêmes régions de leur domaine N-terminal pour leur interaction respective avec MutS et MutS α . Une autre similitude est que cette interaction ne dépend pas de la fixation de l'ATP sur MutL α puisque des mutations dans le domaine ATPasique de MLH1 ou PMS2 n'altèrent pas la formation du complexe MutS α - MutL α (Raschle et al., 2002).

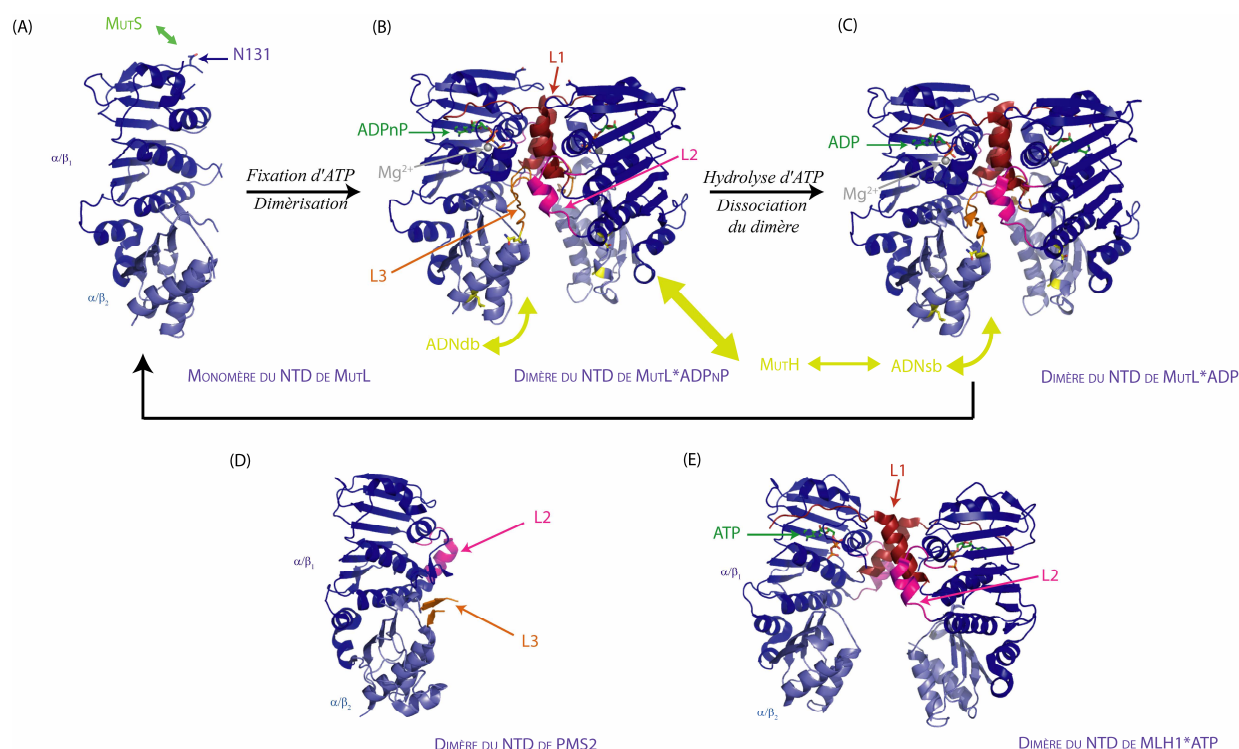


Figure 1 - 8: Structure du domaine N-terminal de MutL d'*E. coli* et de ses homologues (A) Structure cristallographique (code PDB: 1BKN) à 2.90Å du domaine N-terminal d'ecMutL; les deux sous-domaines α/β sont représentés en bleu clair et bleu foncé; la chaîne latérale du résidu N131, impliqué dans l'interaction avec ecMutS, est indiquée (B) Structure cristallographique (code PDB : 1B63) à 1.90Å du domaine N-terminal d'ecMutL lié à une molécule d'ADPNP en présence de Mg^{2+} ; les boucles L1, L2, L3 sont respectivement coloriées en rouge, magenta et orange; les résidus impliqués dans l'interaction avec MutH sont présentés en jaune (C) Structure cristallographique (code PDB : 1B62) à 2.10Å du domaine N-terminal d'ecMutL lié à une molécule d'ADP en présence de Mg^{2+} (D) Structure cristallographique (code PDB : 1H7S) à 1.95Å du domaine N-terminal de PMS2 (E) Structure cristallographique (code PDB : 3NA3) à 2.50Å du domaine N-terminal de MLH1 lié à une molécule d'ATP. Ces représentations montrent la similarité structurale entre ces différents domaines N-terminaux.

Le domaine C-terminal de MutL et de ses homologues

Le domaine C-terminal (ou CTD pour C-terminal domain) de MutL et de ses homologues a d'abord été identifié comme le site de dimérisation principal des protéines MutL bactériennes (Drotschmann et al., 1998) ou eucaryotes (Pang et al., 1997). Au sein du MMR eucaryote, les homologues de MutL forment trois hétérodimères MutL α , MutL β et MutL γ . Une combinaison de double-hybride et de co-IP a permis d'établir que les protéines MLH3, PMS1 et PMS2 partagent un même site d'interaction sur MLH1 compris entre les résidus 492 et 742 du CTD de MLH1. Toutes trois interagissent par leur domaine C-terminal également (résidus 612-674 de PMS2, résidus 1399-1424 de MLH3; résidus 693-728 de PMS1). Ces quatre protéines utilisent des zones similaires de leur région C-terminale pour réaliser ces interactions d'hétérodimérisation (Kondo et al., 2001).

Les régions C-terminales des protéines MutL bactériennes et eucaryotes présentent des séquences peu conservées entre bactéries et eucaryotes (environ 10% d'identité de séquences entre *E. coli* et l'humain). De nombreuses différences fonctionnelles ont été reportées ces dernières années provenant notamment de

l'absence d'endonucléase MutH chez la plupart des organismes en dehors d'un ensemble restreint de bactéries à Gram négatif comme *E. coli*. Il est connu depuis 2005, que l'activité endonucléase essentielle pour le bon fonctionnement du MMR est prise en charge par les protéines MutL chez les organismes ne possédant pas l'endonucléase MutH. Ainsi les homodimères MutL de *B. subtilis*, *N. gonorrhoeae* ou *A. aeolicus* ont une fonction endonucléase portée par leur domaine C-terminal (Duppatta et al., 2009; Iino et al., 2010; Pillon et al., 2010a). Cette activité existe également chez les hétérodimères eucaryotes homologues de MutL. Le passage d'homodimère à hétérodimère tend à faire évoluer chaque partenaire vers plus de spécificité; ainsi l'activité endonucléase est portée par Pms1/PMS2 chez MutL α (Kadyrov et al., 2006; Kadyrov et al., 2007) et Mlh3 chez MutL γ (Nishant et al., 2008). Aucune activité endonucléase n'a été identifiée sur Mlh1 et sur MutL β .

La présence de la fonction endonucléase dans la région C-terminale des homologues de MutL suggère que cette région interagit avec l'ADN au moins au moment de l'incision. Il a été proposé que les domaines N- et C-terminaux venant entourer l'ADN à travers la structuration de la chaîne les reliant (Kosinski et al., 2005; Sacho et al., 2008).

La région C-terminale de Mlh1 est également impliquée dans un certain nombre d'interactions protéine-protéine qui seront détaillées un peu plus loin car ils sont au centre des questions adressées dans cette thèse

iii. Mutations des homologues de MutL associés au syndrome de Lynch et identifiées comme délétères

Comme cela a déjà été précisé, les mutations concernant MLH1 représentent quasiment la moitié des cas de syndrome de Lynch et un tiers de ces mutations sont des mutations ponctuelles avec la substitution d'un seul acide aminé sur la protéine. Cependant la classification de ces variants faux-sens est assez compliquée car d'une part, même si un certain nombre de phénotypes sont communs à tous les syndromes de Lynch, ces mutations ponctuelles sont particulièrement sujettes à l'obtention de symptômes cliniques excessivement larges. D'autre part, les cas cliniques sont peu nombreux et le sont encore moins si l'on ne considère que les variants ponctuels. Ainsi les exemples de comparaison de deux cas cliniques portant une même mutation sont rares (Peltomaki and Vasen, 2004).

Concernant la protéine MLH1, une étude a récemment recensé l'ensemble des variants faux-sens identifiés (Chao et al., 2008): des quelques 200 mutations répertoriées, seules 37 mutations entraînent une pathogénicité avérée. 17 mutations sont qualifiées de neutres car elles sont également trouvées chez des populations ne présentant pas de syndrome de Lynch. L'ensemble des autres mutations, à savoir la très grande majorité, sont qualifiées d'UVs (pour ununclassified variants), c'est-à-dire des variants non classés. L'étude biochimique de ces variants est essentielle pour déterminer l'effet délétère ou non et palier ainsi le peu d'information clinique.

Si l'on ne considère que les mutations clairement identifiées comme délétères, 20 mutations sont localisées dans la région N-terminale de MLH1 (positions E23, P28, D41, S44, D63, G67×3, C77, F80, K84, I107, T117, L155, R182, V185, S247, R265, D304 et H329). 17 mutations sont localisées dans le domaine C-terminal (positions V506, L550, N551, L559, P581, L582, A589, D601, L622, P648×2, P654, R659×2, D737, L749, R755). Aucun variant délétère n'est identifié dans la région charnière reliant les deux domaines N- et C-terminaux.

L'étude biochimique des variants délétères comme celle des UVs de MLH1 est essentielle pour mieux comprendre l'impact des mutations ponctuelles recensées sur l'expression, la stabilité, la structure et la fonction de la protéine. Depuis une dizaine d'années, quelques études ont montré qu'une part importante des mutations ponctuelles sur le gène aboutit à une moindre expression de la protéine ou à une mauvaise localisation cellulaire. Cependant, plusieurs variants correctement exprimés et localisés, présentent des fonctions altérées, que ce soit au niveau du site d'interaction avec MutS α (Plotz et al., 2006), du site ATPase (Guarne et al., 2004) ou du site d'hétérodimérisation (Kosinski et al., 2010).

Il existe également plusieurs études concernant cette fois, les UVs de MLH3 (Korhonen et al., 2008; Ou et al., 2009); toutes concluent que l'implication de cette protéine dans le syndrome de Lynch n'est pas certaine. Les données biochimiques concernant les UVs de PMS2 sont, pour leur part, peu nombreuses à ce jour.

1.2 CONTEXTE DU PROJET: ETUDE STRUCTURALE ET FONCTIONNELLE DE LA REGION C-TERMINALE DES HOMOLOGUES EUCARYOTES DE MUTL

1.2.1 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE MUTL

La première structure de domaine C-terminal de MutL résolue a été la structure du CTD de MutL d'*E. coli* (Guarne et al., 2004). La région cristallisée par Guarné et al. est le domaine compris entre les résidus S432 et D615, c'est-à-dire le domaine C-terminal prédit par Psipred (résidus 439-615) mais un peu plus court que le fragment obtenu par protéolyse ménagée (résidus 351-615). La structure cristallographique résolue à 2.1Å par méthode MAD contient deux monomères dans l'unité asymétrique. Chaque monomère du domaine CTD de MutL se décompose en deux sous-domaines: un domaine appelé initialement «Ex» (résidus 432-474 et 570-615) correspondant aux résidus localisés aux extrémités N- et C-terminales de la structure et un domaine «In» (résidus 475-569) correspondant à la séquence interne du CTD de MutL (**Figure 1 - 9(A)**). Chaque sous-domaine est composé d'un feuillet β

et de trois hélices α auxquels il faut ajouter pour le domaine Ex, l'hélice α_G à l'extrémité C-terminale. Cette dernière hélice forme une protubérance par rapport au domaine auquel elle appartient. Les structures cristallographiques de MutL de *B. subtilis* et *N. gonorrhoeae* résolues plus récemment suggèrent que les deux sous-domaines peuvent pivoter l'un par rapport à l'autre. L'analyse par la méthode des modes normaux, que nous avons réalisée, suggère également des mouvements possibles des sous-domaines du CTD de MutL (**Figure 1 - 9(B)**). Il a été proposé que ces mouvements jouent un rôle dans la régulation des interactions du CTD de MutL avec les protéines partenaires de cette région.

L'interface de dimérisation initialement proposée pour le CTD de MutL par Guarné et al. est différente de celle communément admise aujourd'hui. Guarné et al. ont proposé que cette interface se trouve dans le domaine «In». Cette interface fait intervenir les hélices α_C de chaque monomère et présente une surface enfouie totale de 1000Å². Ce modèle correspond à un dimère en forme de V (**Figure 1 - 9(A)**). L'équipe de P. Friedhoff a alors observé dans l'empilement cristallin du CTD de MutL, une autre interface avec des molécules symétriques. Cette seconde interface qui implique la région «Ex» du CTD de MutL correspond à une surface enfouie totale plus grande de 1800Å² et contient des acides aminés plus conservés dans les séquences des MutL bactériens et eucaryotes que la première interface. Les auteurs de cette étude ont alors introduit par mutagenèse dirigée des cystéines sur chacune des deux interfaces et ont montré que la seconde interface pouvait être pontée beaucoup plus efficacement par cette approche que l'interface initialement proposée. Les structures résolues ultérieurement de *B. subtilis* et *N. gonorrhoeae* ont confirmé les travaux du laboratoire de P. Friedhoff. L'interface de dimérisation est essentiellement composée du feuillet β de ce sous-domaine et de l'hélice α_G qui vient pour chaque monomère faire une accroche sur l'autre monomère (**Figure 1 - 9(C)**). Cette hélice est essentielle pour la dimérisation puisque sa délétion donne une protéine MutL monomérique non fonctionnelle (Kosinski et al., 2005). Ainsi, la surface de dimérisation initialement proposée par Guarné et al. constitue un contact cristallin dans l'empilement du CTD de MutL. Cependant, ce site contient des résidus conservés; cette région a, par la suite, été caractérisé comme important pour l'interaction avec l'endonucléase MutH par des expériences de pontage et des mesures par spectrométrie de masse (Ahrends et al., 2006). Le domaine C-terminal interagit également avec d'autres partenaires : l'hélicase UvrD (Hall et al., 1998), l'endonucléase Vsr (Drotschmann et al., 1998) et les sous unités β -clamp de la polymérase Pol III (Li et al., 2008).

C'est à la vue de ces données que le sous-domaine Ex a ensuite été nommé sous-domaine de dimérisation tandis que le sous-domaine In est appelé sous-domaine régulateur.

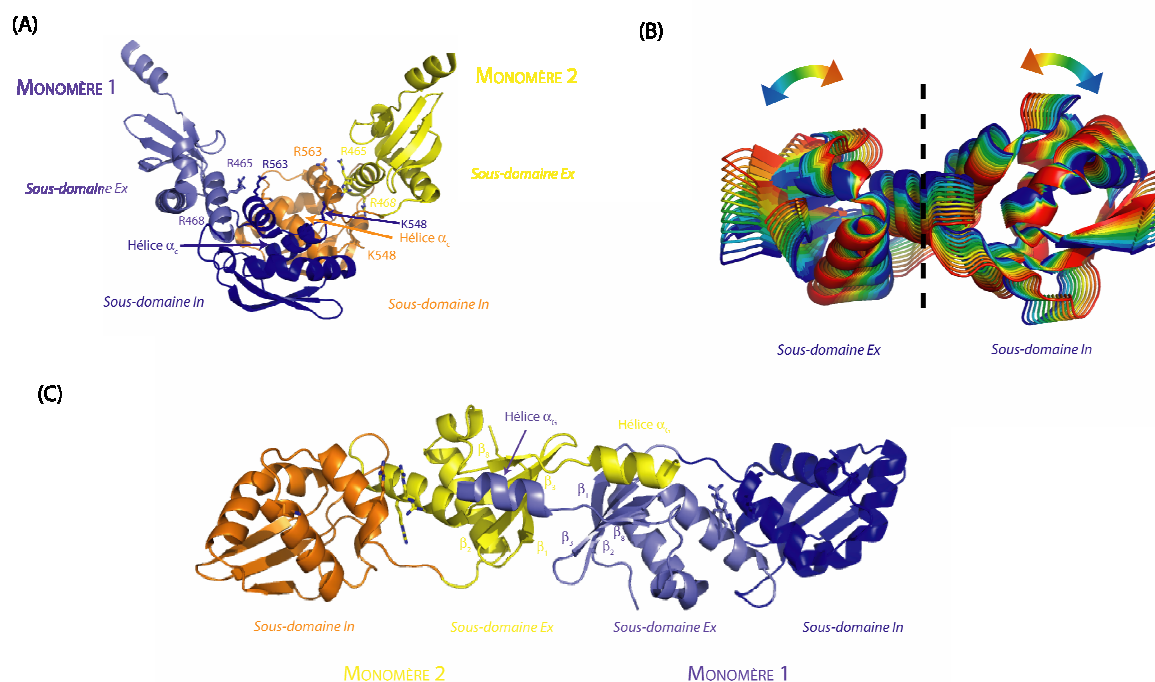


Figure 1 - 9 : Structure du domaine C-terminal de MutL d'*E. coli* (A) Structure cristallographique (code PDB : 1X9Z) à 2.10Å du domaine C-terminal d'*ecMutL* selon le modèle de W. Yang ; les sous-domaines In sont représentés en bleu foncé et orange, les sous-domaines Ex sont représentés en bleu clair et jaune ; les chaînes latérales des résidus identifiés comme importants pour fixer l'ADN dans cette étude, sont détaillées (B) Représentation des onze modèles perturbés induits par le mode normal 7 du domaine C-terminal d'*ecMutL* générés par le programme *ELNémo* (Suhre and Sanejouand, 2004); la perturbation minimale est de -100 (en bleu), celle maximale est de +100 (en rouge) par pas de perturbation de 20 ; l'hélice α_G a été supprimée car les perturbations appliquées la déstructuraient de façon non réaliste (C) Modèle du dimère du domaine C-terminal d'*ecMutL* selon le modèle de J. Kosinski

1.2.2 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE PMS2: ACTIVITE ENDONUCLEASE DU DOMAINE C-TERMINAL

Comme évoqué précédemment, dans le cas des organismes eucaryotes et de beaucoup de bactéries, l'analyse de séquences n'a pas permis de détecter la présence d'homologues de MutH. Pour une partie de ces organismes, il a été découvert une faible activité endonucléase du complexe MutL α . Cette activité catalytique vient palier l'absence d'endonucléase de type MutH.

Deux structures de MutL de bactéries présentant une activité endonucléase, ont été résolues récemment: celle de MutL de *B. subtilis* (Pillon et al., 2010a) et celle de MutL de *N. gonorrhoeae* (Namadurai et al., 2010). Ces structures montrent que le domaine C-terminal de ces protéines a un repliement tridimensionnel proche de celui de MutL d'*E. coli*. Chaque monomère est composé de deux sous-domaines, le domaine de dimérisation et le domaine régulateur; ces domaines sont constitués d'un feuillet β et de 3 hélices α . Ces domaines se superposent bien avec ceux d'*E. coli*. Différentes études proposent que Mlh1-Pms1 de levure ou MLH1-PMS2 humain

possèdent une interface d'hétérodimérisation proche de celle proposée pour le CTD de MutL par Guarné et al. (Cutalo et al., 2006) (**Figure 1 - 10(A)**).

Cependant, de par leur activité endonucléase, les protéines MutL bactériennes de *B. subtilis* et *N. gonorrhoeae* présentent des différences de structure par rapport à MutL d'*E. coli* dans une région située entre le domaine de dimérisation et le domaine régulateur. Cette région adopte une conformation différente liée principalement à une boucle entre l'hélice α_E et le brin β_8 (**Figure 1 - 10(B)**). Ce site est composé de trois motifs, initialement identifiés chez les eucaryotes et très conservés: le 1^{er} motif identifié a pour séquence consensus DQHA(X)₂E(X)₄E (Kadyrov et al., 2006; Kadyrov et al., 2007). La mutation de l'acide aspartique ou du 1^{er} glutamate de ce motif a été effectuée sur PMS2 l'humain, Pms1 de levure ou sur MutL des bactéries *B. subtilis* et *N. gonorrhoeae*. Ces mutations entraînent une perte de la fonction endonucléase de ces protéines. Deux autres motifs très conservés (A/S)C(R/K) et C(P/N)HGRP forment le site catalytique des MutL bactériens. Ces motifs sont retrouvés chez les protéines Pms1/PMS2 et ont été proposés comme faisant partie du site catalytique de ces protéines (Yang, 2006) et (Kosinski et al., 2008) (**Figure 1 - 10(C)**). Ces trois motifs ont également été identifiés chez les bactéries thermophiles *A. aeolicus* (Iino et al., 2010) et *T. thermophilus* (Fukui et al., 2008).

Il est intéressant de noter que ces motifs sont présents dans la séquence de la protéine Mlh3. L'étude génétique chez *S. cerevisiae* de mutants des motifs endonucléases de Mlh3 met en évidence un taux de mutation 6 fois supérieur à celui d'une levure sauvage ainsi qu'un défaut dans la formation des crossover au cours de la méiose (Nishant et al., 2008). Aucun de ces motifs n'a pu être identifié sur la protéine PMS1 (**Figure 1 - 10(C)**).

Comme beaucoup d'endonucléases, les CTD de la famille PMS2 utilisent un ou plusieurs cations divalents comme cofacteurs (Galburt and Stoddard, 2002). Chez *S. cerevisiae* comme *H. sapiens*, la fonction d'endonucléase est activée par Mg²⁺ en présence de MutS α , RFC, PCNA et d'ATP (Kadyrov et al., 2006; Kadyrov et al., 2007). En l'absence de ces protéines, MutL α humaine ou de levure est activée en présence d'ADN homoduplex ou hétéroduplex en présence de Mn²⁺ à 1mM. L'étude de MutL de *N. gonorrhoeae* montre que l'endonucléase est fonctionnelle en présence d'ions Mg²⁺, Mn²⁺ ou Ca²⁺ par ordre décroissant d'efficacité catalytique (Duppatla et al., 2009). MutL de *B. subtilis* n'est fonctionnelle qu'en présence d'ions Mn²⁺ et non de Mg²⁺, Zn²⁺ et Cd²⁺ tandis que MutL de *T. thermophilus* nécessite soit Mn²⁺, soit Ni²⁺ ou encore Co²⁺ mais pas Mg²⁺, Zn²⁺ ou Ca²⁺. Enfin, la protéine d'*A. aeolicus* se comporte comme celle de *T. thermophilus* à la différence près que cette première présente une activité en présence de 5mM de ZnCl₂ (Iino et al., 2010). Ainsi les données biochimiques sur les différentes endonucléases MutL suggèrent la présence de Mg²⁺ ou Mn²⁺ au sein des sites catalytiques des homologues de MutL sans pour autant exclure d'autres possibilités comme Zn²⁺ et Ca²⁺ notamment..

Les structures cristallographiques de MutL à activité endonucléase ont apporté des éléments d'information intéressants sur le site endonucléase. Dans l'une des formes cristallines du CTD de MutL de *B. subtilis*, deux atomes de zinc sont observés au sein du site formé par les trois motifs conservés. Il est à noter que les cristaux correspondant ont été obtenus avec une solution précipitante contenant 0.5mM de ZnCl_2 . Les deux atomes de zinc sont coordonnés par le résidu E468 du 1^{er} motif, formant ainsi un contact avec les deux atomes métalliques et par une cystéine et une histidine de chaque côté (H464 du 1^{er} motif et C573 du 2nd d'une part; H606 et C604 du 3^e motif d'autre part). Le résidu D462 essentiel pour la fonction n'est pas directement impliqué dans la coordination des cations (**Figure 1 - 10(D)**).

Les analyses bioinformatiques du site métallique prédisent un site de fixation du zinc (Kosinski et al., 2008). Comme toutes les structures disponibles présentent des ions Zn^{2+} dans le site catalytique, ces analyses suggèrent que le zinc peut jouer un rôle direct ou même de régulation. Cependant, hormis chez *A. aeolicus*, aucune activité n'a été détectée en présence de zinc. De l'ensemble de ces données, il ressort que les cofacteurs les plus probables sont le magnésium et le manganèse même si le rôle de certains autres cations comme le zinc ne peut être totalement écarté. Pillon et al. (Pillon et al., 2010a) ont observé que l'ajout de 1mM de Zn^{2+} ou de Co^{2+} en complément du manganèse accroît l'activité catalytique de MutL de *B. subtilis*. Ces résultats montrent qu'il est possible que les deux cations métalliques au sein du site endonucléase ne soient pas les mêmes. Il a été proposé qu'un site possède un rôle structural alors que l'autre présente un rôle catalytique. Il est ainsi possible que le second site nécessite la présence du cation Mg^{2+} ou Mn^{2+} en accord avec les données biochimiques.

Le domaine N-terminal des protéines MutL bactériennes ou eucaryotes semble avoir un rôle significatif par rapport à l'activité endonucléase portée par la région C-terminale. Une nouvelle fois, les quelques études existantes montrent des comportements différents. Des expériences de réparation MMR *in vitro* montrent que des mutations dans les domaines ATPasique du complexe MLH1-PMS2 (E34 sur MLH1 ou E41 sur PMS2) donnent des complexes MutL α inactifs pour la réparation d'un mésappariement dirigé en 3' (Kadyrov et al., 2006). Le comportement de MutL de *B. subtilis* confirme l'importance du domaine N-terminal puisque le domaine C-terminal seul ne présente pas d'activité endonucléase (Pillon et al., 2010a). En revanche, les domaines C-terminaux seuls de *N. gonorrhoeae* (Duppatla et al., 2009) et d'*A. aeolicus* (Mauris and Evans, 2009) sont suffisants pour l'incision du brin d'ADN bien que l'activité soit accrue en présence de la région N-terminale (Iino et al., 2010). Une explication proposée est que le domaine N-terminal de MutL participe à la fixation de l'ADN devant être incisé par les protéines MutL (Hall et al., 2003). La fixation de l'ATP sur le domaine N-terminal, activité impliquée dans la liaison à l'ADN, permettrait de stabiliser l'interaction ADN*CTD nécessaire à l'hydrolyse de

la liaison phosphodiester. Cette vision est corroborée par les changements conformationnels, directement observés par microscopie à force atomique sur MutL α de levure (Sacho et al., 2008) et par SAXS sur MutL de *T. maritima* (Kim et al., 2009). L'étude sur MutL d'*A. aeolicus* vient également confirmer ce rôle du NTD puisque l'incubation du NTD et du CTD, purifiés séparément, augmente l'activité endonucléase du CTD sur un plasmide (Iino et al., 2010).

L'équipe d'A. Guarné a montré récemment par des expériences de pontage que le CTD de bsMutL interagit spécifiquement avec le β -clamp *via* un site conservé du sous-domaine régulateur. Ce motif a pour séquence consensus QX Φ (L/I)XP, séquence très proche des autres motifs d'interaction avec le β -clamp (QLXLF) (Pillon et al., 2010b). Ils ont montré par génétique que des mutations sur ce motif entraînent un phénotype mutateur très fort chez *B. subtilis* et fort chez *E. coli*. Il a alors été proposé que cette interaction joue un rôle dans la régulation des deux sites endonucléases présents sur les homodimères bactériens. D'après les structures de CTD de MutL, les deux sites endonucléases de l'homodimère se trouvent sur des faces opposées (Namadurai et al., 2010). En prévention d'une double incision par ces deux sites, un des sites pourrait être inhibé par d'autres protéines du MMR comme le β -clamp, favorisant l'activation du site actif le plus proche de l'ADN à inciser.

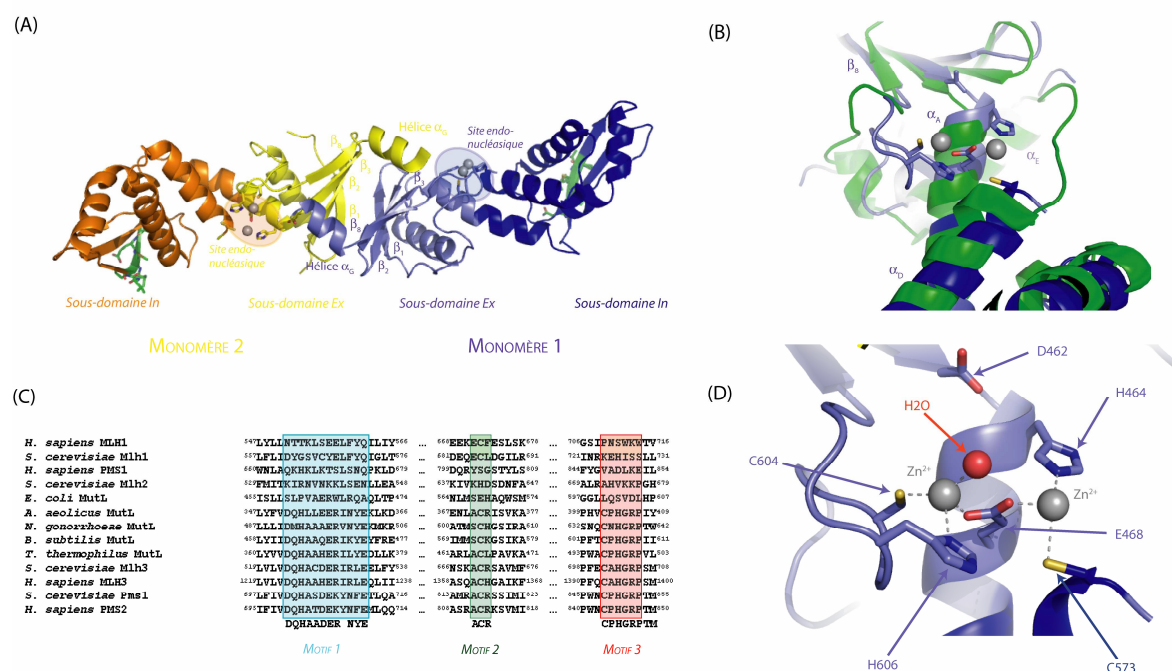


Figure 1 - 10 : Structure du domaine endonucléasique (A) Structure cristallographique (code PDB: 3KDK) à 2.26Å du domaine C-terminal de bsMutL; les domaines In ou régulateur sont représentés en bleu foncé et orange, les domaines Ex ou dimérisation sont représentés en bleu clair et jaune; le domaine d'interaction avec le β -clamp en vert (B) Superposition du domaine endonucléase de bsMutL (en bleu) et du domaine équivalent de ecMutL (en vert) (C) Alignement des motifs impliqués dans le domaine endonucléase pour les organismes où cette activité a été identifiée (D) Détail structural du domaine endonucléase de bsMutL

1.2.3 ETUDE STRUCTURALE ET FONCTIONNELLE DE LA FAMILLE MLH1

Différentes interactions ont été décrites avec la protéine Mlh1 avec des protéines du MMR mais également avec des protéines impliquées dans d'autres mécanismes de réparation, de recombinaison ainsi que dans d'autres processus cellulaires (cf. §1.1.1.*iv*).

Ainsi, une étude de l'interactome de MLH1 humain par purification d'affinité en tandem (TAP-tag) a montré que cette protéine peut être isolée en complexe avec 35 protéines principalement (Cannavo et al., 2007). On retrouve parmi les partenaires identifiés par Tap-Tag, les protéines MSH2 et MSH6 qui interagissent *via* MSH2 et le domaine N-terminal de MLH1. Dans cette étude, sont également retrouvés les partenaires qui forment les hétérodimères avec MLH1, PMS2, MLH3.

Au-delà de ces protéines strictement impliquées dans le MMR, différentes protéines impliquées dans d'autres processus cellulaires ont été identifiées par Tap-Tag. Certaines de ces interactions ont été confirmées par co-IP comme l'interaction entre MLH1 et l'Angiomotine ou KIAA1018/FAN1 (Cannavo et al., 2007; Smogorzewska et al., 2010). En parallèle, des interactions ont également été mises en évidence par d'autres approches avec MLH1 humain ou Mlh1 de levure. On peut citer les interactions de MLH1 avec la protéine de contrôle du cycle cellulaire Rad9 (He et al., 2008) et la glycosylase MED1 (aussi appelée MBD4) (Bellacosa et al., 1999). On ne connaît pas encore les régions d'interactions de tous les partenaires de Mlh1 de levure ou humain. Cinq partenaires ont été identifiés comme ayant au moins un de leur domaine d'interaction sur le domaine C-terminal.

Le premier rôle fonctionnel assigné au domaine C-terminal de Mlh1 est la formation des hétérodimères avec Pms1, Mlh2, Mlh3 chez la levure et PMS2, PMS1, MLH3 chez l'homme.

Un second rôle majeur attribué au domaine C-terminal de Mlh1 est le recrutement de l'exonucléase Exo1. Les études d'interactions entre Mlh1 de levure ou humain ont montré que cette interaction met en jeu le domaine C-terminal de Mlh1 (Schmutte et al., 2001; Tran et al., 2001). Le domaine ATPasique n'est pas nécessaire pour cette interaction mais, tout comme MutL vis-à-vis de MutH et UvrD, ce domaine est possiblement nécessaire pour stimuler l'activité exonucléase d'Exo1 sur le brin néosynthétisé.

Comme précédemment évoqué (cf. §1.1.1.*iii*), un crible en double hybride avec Mlh1 de levure comme appât a permis d'identifier une interaction entre Mlh1 et le facteur de processivité PCNA (Umar et al., 1996) et plus récemment, le groupe d'E. Alani a testé la présence d'un motif PIP sur le domaine C-terminal de Mlh1 de levure, porté par les résidus Q572, L575 et F578 (Lee and Alani, 2006).

La région C-terminale de MLH1 humain (résidus 478-756) est également impliquée dans l'interaction avec le domaine hélicase de la protéine FANCIJ (également connue sous le nom de BACH1/BRIP1) (Peng et al., 2007). Cette dernière, dont certaines mutations sont impliquées dans une prédisposition héréditaire aux cancers du poumon, co-précipite avec MLH1 seul, l'interaction étant cependant renforcée par la présence de PMS2. Le site d'interaction sur la protéine MLH1 n'est pas entièrement connu mais très récemment une mutation ponctuelle L607H, impliquée dans le syndrome de Lynch, a été associée à la perte d'interaction entre les protéines MLH1 et FANCIJ. L'hélicase FANCIJ a été décrite comme essentielle dans la réponse aux dommages de l'ADN de type pontage interbrin et les défauts de progression du cycle cellulaire qui en découlent. Or, comme cela a été évoqué précédemment, la protéine MLH1 aide à reconnaître ce type d'adduit. Le rôle de cette interaction n'est pas bien compris mais il a été observé que l'expression de mutants du site d'interaction de FANCIJ produit une sensibilité aux pontages interbrin.

Une autre protéine impliquée dans le même mécanisme interagit avec le complexe MutL α . En effet, la polymérase translésionnelle Pol η a été décrite récemment comme interagissant avec le CTD de MLH1 (résidus 473-756) par pull down et co-IP (Kanao et al., 2009). Le site d'interaction sur la surface de MLH1 n'est pas connu mais semble différent du site de dimérisation puisque Pol η et PMS2 peuvent interagir simultanément avec la région C-terminale de MLH1. La signification de l'interaction n'est pas bien comprise pour le moment. Il a été proposé que la présence des protéines du MMR dans le processus d'hypermutation somatique (SHM) facilite le recrutement de la polymérase translésionnelle Pol η également présente dans ce processus. Cependant les souris *mlh1*^{-/-} et *pms2*^{-/-} ne présentent pas de phénotype particulier du point de vue SHM. Un des modèles proposés aujourd'hui met en avant le recrutement de Pol η face à la lésion reconnue par le mécanisme MMR (Xie et al., 2010).

Enfin, il a été souligné que la voie MMR présentait des connections probables avec d'autres voies de réparation comme le BER ou la recombinaison homologue. La protéine, par son domaine C-terminal, interagit avec deux protéines de ces voies de réparation. Ainsi, la protéine Mlh1 de levure interagit en double hybride, co-IP et Far Western, avec l'hélicase Sgs1. MLH1 humain interagit de son côté avec un des homologues de Sgs1 chez l'humain, l'hélicase BLM de la famille des hélicases RecQ (Langland et al., 2001; Pedrazzi et al., 2001). L'étude d'extraits cellulaires de patients présentant un syndrome de Bloom, montre que l'hélicase BLM ne semble pas jouer de rôle dans le MMR. Le rôle de l'interaction BLM*MLH1 n'est pas élucidé aujourd'hui.

Une interaction entre le domaine C-terminal de Mlh1 et la glycosylase Ntg2 a été identifiée dans le laboratoire de S. Boiteux au cours d'un crible de double hybride avec comme appât trois glycosylases de levure, Ogg1, Ntg2 et Ntg1 (Gellon et al.,

2002). Cette interaction a été confirmée par GST-pulldown. Cette glycosylase de *S. cerevisiae* fait partie des glycosylases de type endonucléase III qui possèdent une large spécificité de substrat puisqu'elles interviennent aussi bien sur les lésions des bases puriques oxydées que sur celles des bases pyrimidiques oxydées. Des mesures de l'activité enzymatique de Ntg2 ont montré que Mlh1 ne stimulait pas l'activité glycosylase de l'enzyme. Du côté du MMR, la délétion du gène de Ntg2 n'a pas d'effet mutateur significatif. La surexpression de la protéine présente en revanche un léger effet mutateur. La protéine Ntg2 n'a pas d'homologue direct chez l'homme. Une interaction a été mise en évidence chez l'homme entre MLH1 humain et une glycosylase. Il s'agit de l'interaction avec la glycosylase humaine MED1, également impliquée dans le BER (Bellacosa et al., 1999). De façon similaire aux résultats obtenus pour l'interaction Mlh1*Ntg2, la protéine MLH1 humaine ne stimule pas l'activité de MED1 et l'interaction avec MED1 n'est pas indispensable pour le bon fonctionnement du MMR.

Le domaine d'interaction sur Ntg2 a été localisé par double hybride en position N-terminale de la protéine (région 15 à 40) en dehors du domaine glycosylase (région 150 à 335). Par mutagenèse dirigée, le motif d'interaction de Ntg2 a été identifié comme constitué de la séquence ₂₃RSKYF. Une séquence similaire a été retrouvée sur deux autres partenaires de la région C-terminale de Mlh1 de levure, Exo1 (₄₄₄RSKFF) et Sgs1 (₁₃₈₂RSKFF). Ces motifs ont été montrés comme essentiels pour l'interaction avec Mlh1. Ces travaux ont permis au laboratoire de S. Boiteux de proposer l'existence d'un motif MIP (pour Mlh1 Interacting Protein) de séquence (R/K)S(R/K)(Y/F)F (Gellon et al., 2002).

1.3 OBJECTIFS DU PROJET: CARACTERISATION BIOPHYSIQUE ET STRUCTURALE DE LA REGION C-TERMINALE DE MUTL α

1.3.1 CARACTERISER LES SITES D'INTERACTIONS DU DOMAINE C-TERMINAL DE MUTL α

Comme évoqué dans cette introduction, la protéine Mlh1 est un acteur essentiel de la voie de réparation MMR de par ses interactions nombreuses au sein du MMR avec les protéines Msh2, Exo1, Pms1, Mlh2, Mlh3 et PcnA chez la levure. Elle est également en interactions avec de nombreux acteurs d'autres voies cellulaires. Mlh1 est ainsi au cœur d'un réseau d'interactions (**Figure 1 - 11**). Le premier volet de ce travail de thèse a porté sur la caractérisation du site d'interaction de Mlh1 de levure avec les protéines Exo1, Ntg2 et Sgs1 qui partagent un motif commun d'interaction avec Mlh1, le motif MIP. Ces premiers travaux, menés en collaboration avec S. Boiteux¹, avaient comme objectifs de répondre aux questions suivantes:

¹ Laboratoire de Radiobiologie de l'ADN, iRCM, CEA-CNRS, Fontenay-aux-Roses

- Les trois protéines de levure Exo1, Ntg2 et Sgs1 qui utilisent un même motif MIP pour interagir avec Mlh1 interagissent-elles bien avec un seul et même site à la surface de Mlh1?
- Ce site se superpose-t-il avec le site d'hétérodimérisation ou correspond-il à un second site différent du site d'hétérodimérisation?
- Quels sont les acides aminés de Mlh1 impliqués dans l'interaction avec les protéines contenant un motif MIP?
- Le site de Mlh1 impliqué dans les interactions avec Exo1, Ntg2 et Sgs1 est-il conservé chez l'homme et est-il impliqué dans les interactions de MLH1 avec les protéines EXO1 et BLM humaines qui contiennent un motif MIP?
- Des mutants du site d'interaction de Mlh1 avec les protéines MIP ont-ils un phénotype particulier au niveau du MMR ou d'autres processus cellulaires?

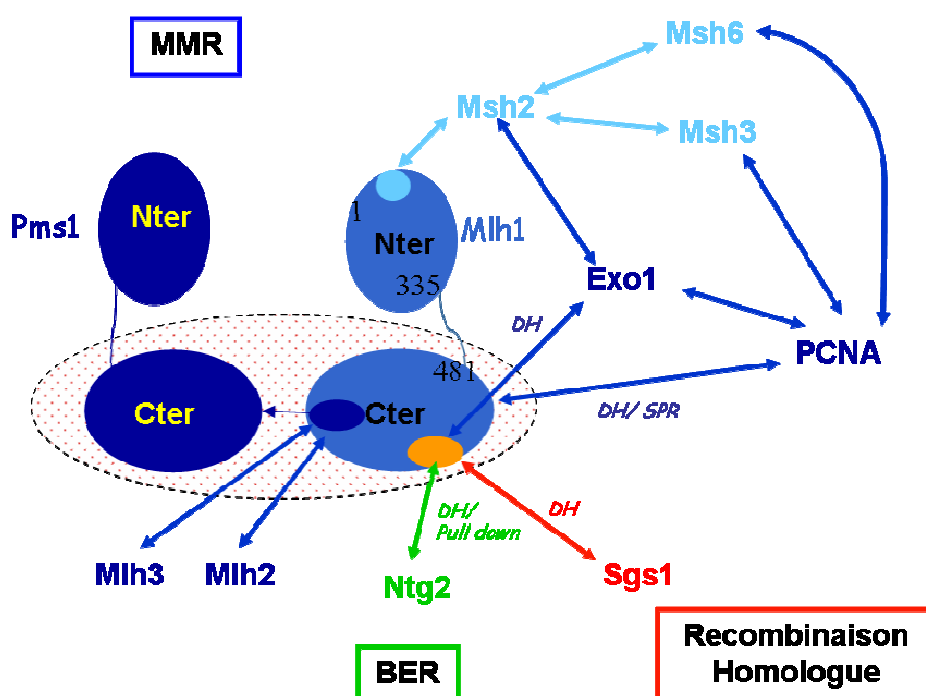


Figure 1 - 11 : Représentation schématique de l'hétérodimère Mlh1*Pms1 et des interactions de la protéine Mlh1 de *S. cerevisiae* avec ces partenaires

1.3.2 RESOLUTION DE LA 1^{ERE} STRUCTURE EUCARYOTE DU DOMAINE C-TERMINAL DE MUTL α

A l'issue de la première partie de mon projet, nous avons choisi de concentrer nos efforts sur la détermination de la structure cristallographique de la région C-terminale de Mlh1 de levure en complexe avec un peptide contenant le motif MIP et/ou avec la région C-terminale de Pms1. Cette étude structurale avait comme objectif de répondre aux questions suivantes:

- Quelle est la similarité structurale des régions C-terminales de MutL α avec les protéines MutL de bactéries possédant une activité endonucléase comme *B. subtilis*

et *N. gonorrhoeae*) ou celles n'en possédant pas comme *E. coli*? Le repliement observé pour ces protéines bactériennes est-il le même que celui observé chez les organismes eucaryotes?

- Quelle est la région de Mlh1 et Pms1 impliquée dans l'hétérodimérisation? Le mode d'hétérodimérisation est-il en tout point semblable au mode d'homodimérisation observé chez les MutL bactériens? Les données structurales suggèrent-elles des propriétés différentes des modes d'interaction de Mlh1 avec les trois sous-unités Pms1, Mlh2 et Mlh3?

- Les acides aminés identifiés, lors de la première partie du travail de thèse, comme essentiels pour l'interaction avec les protéines contenant un motif MIP forment-ils un site d'interaction contigu? Quel est le mode d'interaction du motif MIP avec la protéine Mlh1?

- Quelle est la structure du site endonucléase de Pms1 ? Ce site est-il similaire d'un point de vue structural aux sites endonucléases bactériens précédemment décrits? Quels cations observe-t-on dans le site endonucléase de Pms1?

- La structure de la région C-terminale d'un complexe Mlh1*Pms1 eucaryote permet-elle de localiser les mutations de MLH1 humain associées au syndrome de Lynch? La structure permet-elle notamment d'évaluer l'impact des mutations associées à ce syndrome sur la stabilité, le repliement ou les interactions de MLH1?

Le manuscrit de thèse présente dans la suite, une section «Matériel et méthode» puis une partie «Résultats-Discussion», contenant trois grandes sections : (i) la caractérisation du site d'interaction de Mlh1 de levure et humain avec les protéines contenant un motif MIP. Cette partie a fait l'objet d'une publication en 2009 (Dherin et al., 2009); (ii) la seconde section des résultats présente les différentes étapes qui ont permis de résoudre la structure cristallographique de la région C-terminale de MutL α ; (iii) enfin la dernière section des résultats présente cette structure cristallographique. Les résultats de cette partie font l'objet d'un manuscrit en préparation. Une dernière partie «Conclusion-Perspectives» reprend les résultats principaux de cette thèse et présente les travaux qui nous semblent importants pour la suite du projet.

|

2. Matériel et Méthodes

2.1 EXPRESSION ET PURIFICATION DES PROTEINES

2.1.1 CONSTRUCTION DES DOMAINES

Les protéines Mlh1 et Pms1 sont composées d'un domaine N-terminal et d'un domaine C-terminal reliés par une région charnière prédite comme déstructurée. Les régions C-terminales sont prédites par Jpred, pour débiter aux résidus R498 et Y499 pour Mlh1 de *S. cerevisiae* et *H. sapiens* respectivement et au résidu E651 pour Pms1 de *S. cerevisiae*.

Afin de s'assurer de ne pas tronquer les domaines C-terminaux, les séquences 486-756 de hsMLH1 et 485-769 de scMlh1 sont clonés dans le vecteur pETM-30 {Kan^R} (EMBL) portant une étiquette 6×His-GST de *Schistosoma japonicum*. De même, la séquence 635-873 de scPms1 est clonée dans le vecteur pEXP {Amp^R} (Clontech) portant une étiquette 6×His-MBP d'*Escherichia coli* (Braud et al., 2005). Ces domaines C-terminaux "élargis" correspondent également aux fragments C-terminaux de Mlh1 et Pms1 identifiés dans les cribles de double hybride.

Afin de réduire la taille de la partie déstructurée en début de domaine C-terminal de Mlh1, une seconde construction de scMlh1 (nommée v2, 503 à 769) a été clonée. Les limites de cette seconde construction sont déduites des conclusions de la protéolyse ménagée décrite dans les résultats. En mélangeant 1µl de trypsine (Sigma-Aldrich® - ref T8003) à 100µg/ml à 10ml du domaine C-terminal de scMlh1 (485-769) à 1mg/ml dans le tampon {20mM Tris pH8, 300mM NaCl, 10% glycérol}, on a obtenu une protéine plus courte à l'extrémité N-terminale. La protéine protéolysée a été purifiée par colonne d'exclusion de taille (HiPrep Sephacryl S200 26/60 -ref 17-1195-01) et analysée par séquençage N-terminal et spectrométrie de masse (MALDI-TOF). Le domaine C-terminal de scMlh1 après protéolyse ménagée commence à l'acide aminé 503. La construction 503-769 de scMlh1 a été clonée dans le vecteur pETM-30.

2.1.2 EXPRESSION ET PURIFICATION DES REGIONS C-TERMINALES DE SCMLH1, HSMLH1 ET SCMLH1*SCPMS1

*i. Co-expression et co-purification du complexe Cter de scMlh1*Cter de scPms1*

Co-expression du complexe en milieu riche (LB)

Le complexe d'intérêt est composé des régions C-terminales de scMlh1 (485-769) et scPms1 (635-873). Les deux protéines sont co-exprimées par transformation des cellules par les deux plasmides l'un à la suite de l'autre.

La surexpression du complexe recombinant est réalisée en souche bactérienne BL21(DE3)Star de *Escherichia coli*. Les bactéries préalablement rendues compétentes sont d'abord transformées avec le plasmide pETM-30-GST-scMlh1 et ensemencées en milieu LB additionné de kanamycine.

Ces bactéries sont ensuite rendues chimio-compétentes selon le protocole décrit en ANNEXE1ANNEXE

ANNEXE 1:

- culture de 100 ml de milieu LB à 37°C ensemencé avec une colonie préalablement obtenue. Lorsqu'on mesure une $DO_{600nm} = 0.5$, les cellules sont récoltées par centrifugation (20 minutes à 2500g et à 4°C);
- reprise du culot dans 33 ml de tampon de transformation glacé puis centrifugation (10 minutes à 2500g et à 4°C);
- resuspension du culot dans 8 ml de tampon de transformation glacé {10mM NaHepes pH6.7, 15mM $CaCl_2$, 250mM KCl, 55mM $MnCl_2$ } ;
- ajout goutte à goutte de DMSO à la concentration finale de 7% (v/v). En effet, en présence de certains cations métalliques tels que Mn^{2+} , Ca^{2+} (contenus dans le tampon de transformation) et de certaines molécules organiques telles que le DMSO et le PEG, les parois et membranes bactériennes deviennent perméables par formation de pores autorisant la pénétration de l'ADN (Hanahan, 1983);
- les cellules sont aliquotées et conservées à -80°C après trempage dans l'azote liquide. Ces cellules sont utilisées jusqu'à 3 mois d'âge.

Après transformation par le plasmide pEXP-MBP-scPms1, ces cellules sont cultivées à 37°C en milieu LB additionné des antibiotiques kanamycine (pour le pETM-30-Mlh1) et ampicilline (pour le pEXP-Pms1). Lorsqu'on mesure une $DO_{600nm}=1.2$, la surexpression est induite par ajout d'IPTG à la concentration finale de 0.5mM. La phase d'induction se poursuit à 20°C sur la nuit. Les bactéries sont récoltées et lavées avec du PBS et centrifugées à 4000g, 4°C. A l'issue de la centrifugation les culots bactériens secs sont lysés ou congelés dans l'azote liquide et conservés à -80°C.

Co-expression du complexe en milieu minimum

Le milieu minimum est utilisé pour substituer les méthionines du complexe Mlh1*Pms1 par des sélénométhionines

A partir du produit de transformation liquide, une première préculture est menée en milieu LB, pendant 18 heures, comme précédemment. Cette préculture sert à inoculer le milieu minimum enrichi en sélénométhionines avec les deux antibiotiques appropriés. Le contenu du milieu minimum est présenté en ANNEXE 2.

Co-purification du complexe

Après resuspension dans le tampon de lyse {PBS, 1mM PMSF}, les culots cellulaires contenant les deux protéines de fusion sont lysés:

- d'abord par voie chimique: ajout de lysozyme (Euromedex, Ref 5934-B) à 100mg/l_{culture} pendant 30min;
- puis par sonication: cycle de 1min avec une amplitude allant de 25 à 55% par pas de 10%.

Après clarification par ultracentrifugation et ajout de benzonase (Sigma, E1014-5Ku), le surnageant est purifié en quatre étapes (**Figure 2 - 1**):

- Etape de capture → purification sur résine d'affinité Dextrin-Sepharose (GE Healthcare, ref 28-9355-97) préalablement équilibrée avec le tampon de lyse. La dextrine (qui est un polymère de maltose) permet de fixer la protéine de fusion MBP-Pms1 complexée avec la protéine GST-Mlh1. Après lavage de la résine, les protéines portant l'étiquette MBP sont éluées par ajout de 10mM de maltose au tampon de lyse
- Coupure à la protéase TEV → Les protéines de fusion sont diluées si nécessaire pour être à une concentration inférieure ou égale à 1mg/ml car dans notre cas, des concentrations trop fortes en protéine de fusion s'accompagnent d'une précipitation. On ajoute 1% (m/m) (pour masse/masse) de protéase TEV à l'éluat. La TEV est produite au laboratoire avec une étiquette 6×His (plasmide pET9α de G. Stier - EMBL). Les protéines de fusion sont incubées sur la nuit à 4°C sur un agitateur orbital. Ceci permet de cliver les sites TEV entre les protéines 6×His-MBP et 6×His-GST et respectivement scPms1 et scMlh1.
- Etape de purification sur colonne de Nickel → purification sur colonne d'affinité HisTrap HP (GE Healthcare, ref 17-5248-01, 5ml) préalablement équilibrée avec le tampon de lyse HT {20mM Na₂HPO₄ pH7.4, 500mM NaCl, 20mM Imidazole}. Les protéines portant le peptide hexahistidine, à savoir les protéines MBP, GST ainsi que la protéase TEV, se fixent sur la résine chélatant les cations Ni²⁺. Le complexe d'intérêt est récupéré dans l'éluat direct (FT ou Flow through en anglais);
- Etape de polissage → Deux méthodes ont été utilisées. (i) passage sur colonne d'échange anionique (ResourceTM Q - GE Healthcare -ref 17-1177-01) préalablement équilibrée avec le tampon RQ {20mM Na₂HPO₄ pH7.4, 50mM NaCl, 10mM β-mercaptoéthanol} (ii) passage sur colonne d'interaction hydrophobe (ResourceTM Phe - GE Healthcare-ref 17-1186-01) préalablement équilibrée avec le tampon RP {50mM Na₂HPO₄ pH7.0, 1.5M (NH₄)₂SO₄, 10mM β-mercaptoéthanol}. Sur chromatographie

d'échange anionique, le complexe est fixé à un pH au-dessus de son pI théorique en faible milieu salin et s'élue sur un gradient allant de 0.05 à 1M NaCl sur 20 volumes de colonne. Sur chromatographie d'interactions hydrophobes, le complexe est fixé à forte force ionique et élué sur un gradient allant de 1.5 à 0M $(\text{NH}_4)_2\text{SO}_4$ sur 20 volumes de colonne.

Les rendements obtenus avec ce protocole sont présentés dans la partie résultat. Ce protocole permet d'obtenir le complexe des régions C-terminales de Mlh*Pms1 purifié à homogénéité et en quantité suffisante pour des essais de cristallogénèse.

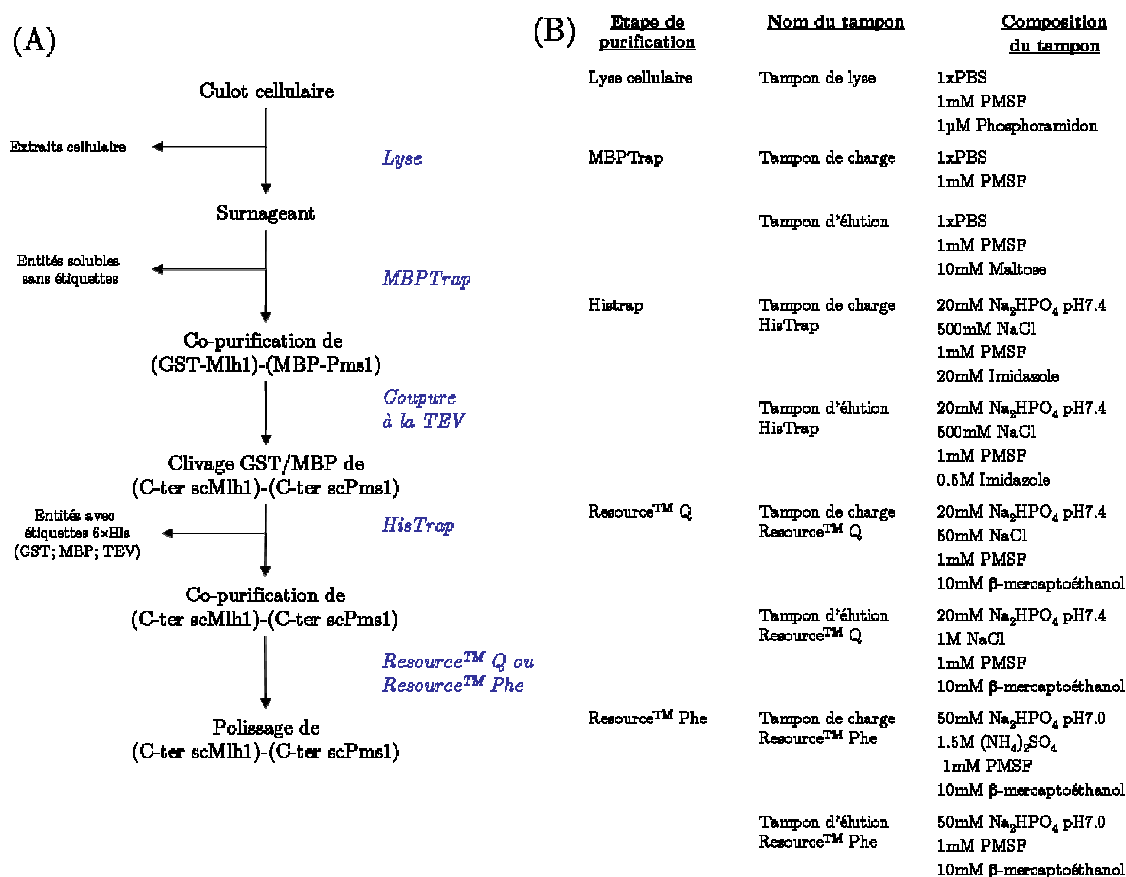


Figure 2 - 1: Purification de la région C-terminale de scMutLα (A) Résumé des étapes de purification (B) Tampons utilisés à chaque étape de purification

ii. Expression et Purification du domaine C-terminal de scMlh1 et de hsMLH1

Expression du domaine C-terminal de scMlh1 et de hsMLH1

L'expression se fait dans des conditions similaires par rapport au complexe (ANNEXE 1). Nous utilisons des bactéries BL21(DE3)Star préalablement rendues compétentes et transformées avec le plasmide pETM-30-GST-scMlh1 dans le cas de la protéine de levure ou pETM-30-GST-hsMLH1 dans le cas de la protéine humaine.

Les deux protéines sont purifiées selon le même protocole.

Purification du domaine C-terminal de scMlh1 et de hsMLH1

Après resuspension dans le tampon de lyse {50mM Na₂HPO₄ pH8, 500mM NaCl, 5% Glycérol, 1% Triton X100, 1mM EDTA, 0.5mM PMSF, 10mM β-mercaptoéthanol}, les culots cellulaires contenant la protéine de fusion sont lysés:

- d'abord par voie chimique: ajout de lysozyme à 100mg/l_{culture} pendant 30min;
- puis par sonication: cycle de 1min avec une amplitude allant de 25 à 55% par pas de 10%.

Après clarification par ultracentrifugation et ajout de benzonase, le surnageant est purifié en quatre étapes (**Figure 2 - 2**):

- Etape de capture → purification sur résine d'affinité GSTrap HP (GE Healthcare, ref 17-5281-01) préalablement équilibrée avec le tampon de charge GT {50mM Na₂HPO₄ pH8, 500mM NaCl, 10mM β-mercaptoéthanol}. Le glutathion (tripeptide composé d'un acide glutamique, d'une cystéine et d'une glycine) permet de fixer la protéine de fusion GST-Mlh1. Après lavage de la résine, les protéines portant l'étiquette 6×His-GST sont éluées par ajout de 10mM de glutathion réduit au tampon de lyse;
- Coupure à la protéase TEV → après dilution de la fraction vers 1mg/ml (pour minimiser la précipitation), on ajoute 1% (m/m) de protéase TEV à l'éluat et on incube sur la nuit à 4°C sur agitateur orbital. Ceci permet de cliver l'étiquette 6×His-GST de la région C-terminale de Mlh1;
- Etape de purification sur colonne de Nickel → purification sur colonne d'affinité HisTrap HP {GE Healthcare} préalablement équilibrée avec le tampon de lyse HisTrap {50mM Na₂HPO₄ pH8, 500mM NaCl, 10mM β-mercaptoéthanol, 20mM Imidazole}. Les protéines portant le peptide hexahistidine, à savoir la protéine GST et la protéase TEV, se fixent sur la résine chélatant les cations Ni²⁺. Le complexe d'intérêt est récupéré dans le "passe-tout-droit";
- Etape de polissage → passage sur chromatographie d'exclusion de taille (ou filtration sur gel) HiPrep Sephacryl S-200 26/60 (GE Healthcare, ref 17-1195-01) préalablement calibrée puis équilibrée avec le tampon GF {50mM Na₂HPO₄ pH7.4, 300mM NaCl, 10mM β-mercaptoéthanol}. Les protéines sont séparées suivant leur volume hydrodynamique.

Les rendements de purification obtenus sont présentés dans la partie résultat. Ce protocole permet d'obtenir les régions C-terminales de scMlh1 et hsMLH1 purifiées à homogénéité et en quantité suffisante pour des essais de cristallogénèse ou des analyses par calorimétrie.

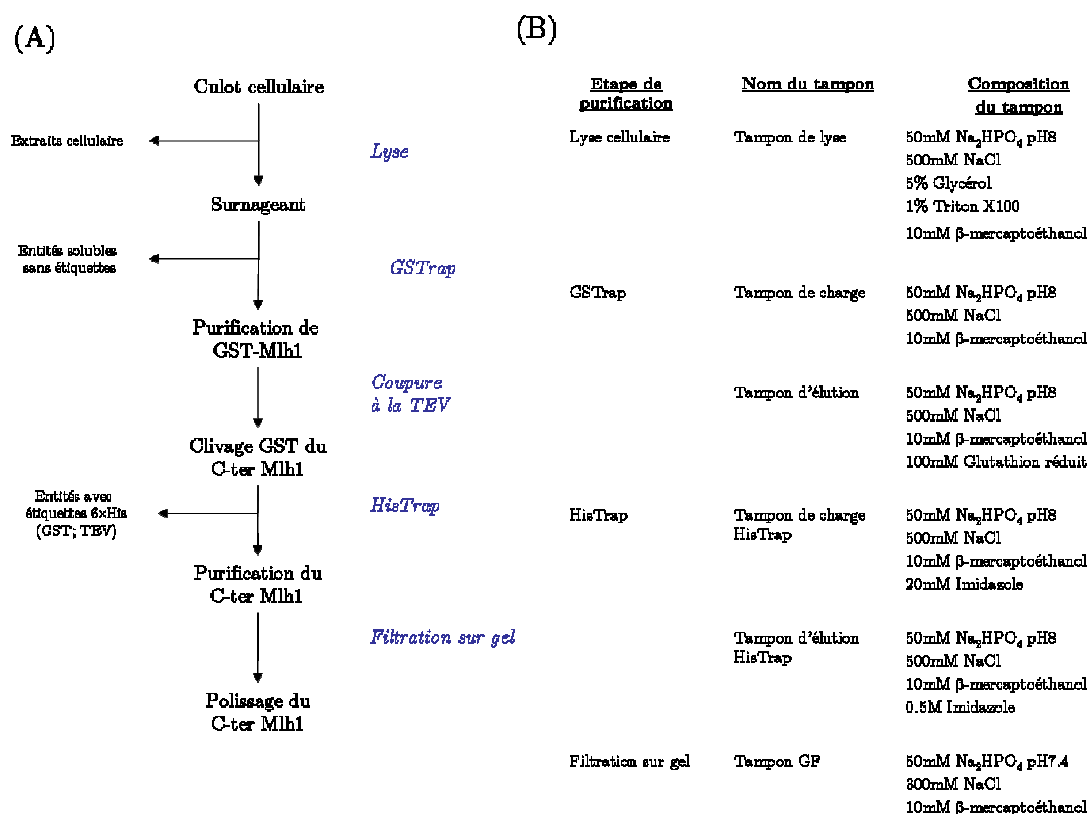


Figure 2 - 2: Purification de la région C-terminale de scMlh1 et de hsMLH1 (A) Résumé des étapes de purification (B) Tampons utilisés à chaque étape de purification

2.2 CARACTERISATION BIOPHYSIQUE DES PROTEINES PURIFIEES

2.2.1 CARACTERISATION DU REPLIEMENT PAR DICHROÏSME CIRCULAIRE

i. Origine du dichroïsme circulaire et utilité pour les protéines

Le dichroïsme circulaire (Circular Dichroism ou CD en anglais) est la mesure de la différence d'absorbance pour des ondes polarisées circulaires droite et gauche (dans la gamme de 170-700 nm). Ce signal est donc observé lorsque les molécules étudiées sont chirales.

Le chromophore le plus abondant dans les peptides et les protéines est la liaison peptidique, qui présente deux bandes d'absorption dans l'UV : une transition $\pi \rightarrow \pi^*$ à 190nm et une transition $n \rightarrow \pi^*$ à 220 nm (Kelly et al., 2005). Les différentes formes de structures secondaires conduisent à des spectres CD dans l'UV lointain, spécifiques et caractéristiques de celles-là. Des protéines qui contiennent surtout des hélices α présentent des bandes fortes à 208 et 222 nm, qui sont absentes des spectres CD des protéines qui contiennent surtout des feuillets β . Ces dernières sont caractérisées par un minimum simple dans le spectre dont la position varie dans la gamme 210-230 nm (Figure 2 - 3).

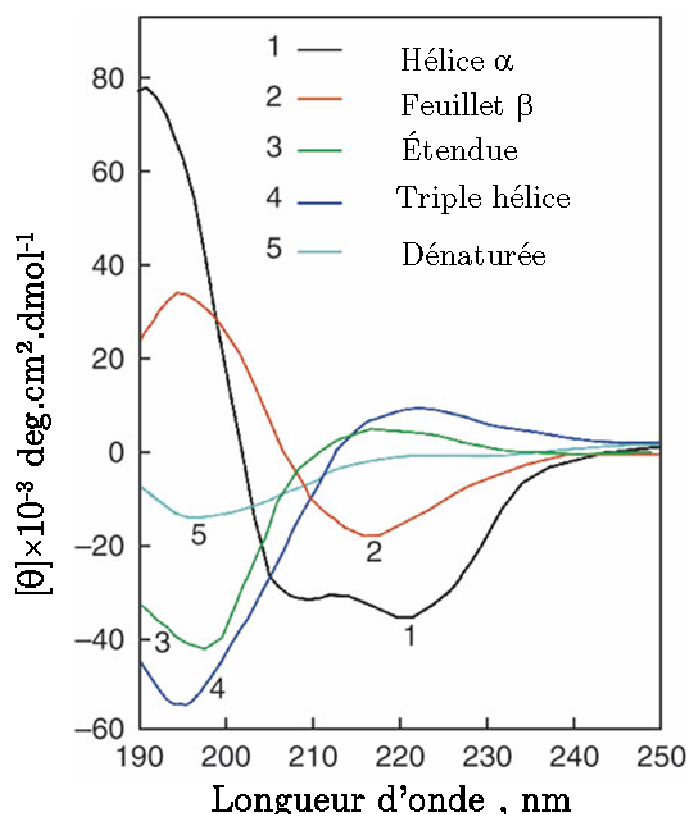


Figure 2 - 3: Exemples de spectres de dichroïsme circulaire pour 1) une hélice α (poly-L-lysine; pH11.1); 2) un feuillet β (poly-L-lysine; pH11.1); 3) une structure étendue (poly-L-lysine; pH5.7); 4) la triple hélice de collagène; 5) une structure dénaturée (collagène dénaturé) (d'après (Greenfield, 2006))

ii. Préparation de l'échantillon et déroulement de la mesure

Le chlorure de sodium absorbe fortement les UV, essentiellement en dessous de 200nm. Afin d'avoir un tampon ayant le moins d'absorbance dans les UV lointains, les échantillons sont dialysés 3 fois pendant 2 heures dans le tampon {50mM Na_2HPO_4 pH7.4, 75mM NaF, 10mM β -mercaptoéthanol} car le fluorure de sodium est optiquement transparent pour les longueurs d'ondes utilisées.

Les échantillons sont ensuite centrifugés 10min à 10 000g et filtrés sur un filtre de 0.1 μm (Millipore); les concentrations sont déterminées par absorption à 280nm sur un Nanodrop (ND1000, LabTech).

Les mesures de CD ont été faites à une concentration comprise entre 4 et 10 μM dans les cellules de 1mm de parcours optique et à une température de 20°C; les spectres sont effectués entre 195 et 260 nm à une vitesse de 50 nm/min sur un spectrophotomètre J-810 (Jasco Corp.) à l'exception des spectres entre la région C-terminale de hMLH1 et des peptides effectués sur un spectrophotomètre J-710 (Jasco Corp.). Chaque spectre est mesuré dix fois pour augmenter le rapport signal/bruit.

Les dénaturations thermiques ont été faites sur les mêmes échantillons entre 5 °C et 95 °C à la longueur d'onde de 222nm par pas de 1°C/min.

iii. Traitement des données

Le signal mesuré est l'ellipticité $[\theta]$ en millidegrés correspondant à l'arc-tangente du rapport du demi-grand axe sur le demi-petit axe de la lumière polarisée elliptiquement (Woldbye and Bagger, 1963). Cette grandeur est reliée au signal CD ($\Delta\epsilon=\epsilon_g-\epsilon_d$) par:

$$[\theta](mdeg) = 10^3 \times \frac{2.303}{4} (A_g - A_d) \frac{180}{\pi} \quad (1)$$

avec A_g et A_d (sans unité), les absorbances respectivement gauche et droite et reliées au signal CD par la loi de Beer-Lambert: $A=\epsilon.l.c$ avec l , la longueur du chemin optique et c , la concentration de l'échantillon.

Les facteurs $10^3, 2.303, \frac{180}{\pi}$ proviennent respectivement du passage de degré en millidegrés, de logarithme base 10 à logarithme népérien, de radian à degré.

Afin de pouvoir comparer les spectres entre eux, le spectre CD est converti en ellipticité molaire résiduelle $[\theta]_{m,res}$:

$$[\theta]_{m,res} (deg \cdot cm^2 \cdot dmol^{-1} \cdot res^{-1}) = 0.1 \times \frac{[\theta]}{l \times c \times n} \quad (2)$$

avec $[\theta]$ (mdeg), l'ellipticité mesurée;

l (cm), la longueur du chemin optique;

c (M), la concentration de l'échantillon.

n , nombre d'acides aminés

L'estimation du degré de structure secondaire est calculé à partir de l'algorithme neuronal K2D (Andrade et al., 1993).

Le dichroïsme circulaire est très utile pour comparer le contenu en structure secondaire d'une protéine sauvage et de variants de cette protéine. Elle permet également dans certains cas d'observer des changements de conformation ou une stabilisation d'une protéine lors de l'interaction avec un peptide ou un ligand. D'un point de vue pratique, elle consomme très peu de matériel mais reste très sensible à l'état d'agrégation (surtout que l'on doit travailler en milieu peu salin), d'où l'intérêt de faire en parallèle l'étude par diffusion de lumière.

2.2.2 CARACTERISATION DE LA POLYDISPERSITE PAR DIFFUSION DYNAMIQUE DE LA LUMIERE

i. Description de la méthode: du mouvement brownien à la distribution de tailles

L'analyse en diffusion dynamique de la lumière apporte des renseignements sur la taille des particules par corrélation entre le mouvement brownien (agitation thermique) et le coefficient de diffusion (D).

Lorsqu'un échantillon de protéines est parcouru par un faisceau de lumière polarisée, une faible partie de ce faisceau est diffusée. Pour une macromolécule de dimension inférieure à la longueur d'onde λ de la source, l'intensité diffusée s'exprime d'après la formule de diffusion de Rayleigh (Cao, 2003).

Au cours du temps, les macromolécules ont un mouvement aléatoire dû aux chocs avec les autres particules (solvant ou protéine). Or d'après la relation de Stokes-Einstein, plus le rayon des macromolécules étudiées est grand, plus leur mouvement est faible (et donc plus leur coefficient de diffusion est faible également):

$$R_H = \frac{k_B T}{6\pi\eta D} \quad (3)$$

où R_H (m), le rayon hydrodynamique apparent des macromolécules;

k_B , le coefficient de Boltzmann ($1.38 \times 10^{-23} \text{ J.K}^{-1}$);

T (K), la température;

D ($\text{m}^2.\text{s}^{-1}$), le coefficient de diffusion;

η (Pa.s), la viscosité du milieu de dilution.

Ces macromolécules étant les sources qui diffusent la lumière, ce mouvement brownien fait varier en fonction du temps l'intensité diffusée $I(t)$ par un volume donné. Cette variation est exprimée par le produit de corrélation $G(t)$:

$$G(t) = \langle I(t) \times I(t + \delta t) \rangle \quad (4)$$

Bien que le mouvement brownien soit aléatoire, il est possible de définir une fonction de corrélation à ce produit $G(t)$ si la variation de temps δt est très petite (de l'ordre de la nanoseconde):

$$G(t) = A \exp(-2Dq^2 \times \delta t) + B \quad (5)$$

avec A , une constante de l'appareil;

B , la ligne de base (théoriquement égale à 1);

D ($\text{m}^2.\text{s}^{-1}$), le coefficient de diffusion;

q (m^{-1}), le vecteur de diffusion égal à $\frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right)$ où θ l'angle de diffusion et n

l'indice de réfraction du milieu.

ii. Description de l'appareil et déroulement de la mesure de distribution de tailles

Les mesures sont réalisées avec un appareil Zetasizer Nano S modèle 1500 (Malvern Instrument Ltd.) équipé d'un laser produisant une source lumineuse verte ($\lambda=532\text{nm}$, puissance=50mW) et relié à un détecteur de photon et un corrélateur. Le dispositif de diffusion de lumière enregistre le signal en rétrodiffusion à un angle $\theta=173^\circ$ par rapport au faisceau incident. La gamme de tailles mesurable est comprise entre 0.6nm et 6 μm .

Les échantillons de protéines sont amenés à une concentration proche de 1mg/ml. Ils sont ensuite filtrés sur filtre de 0.1 μm (Millipore) et centrifugés 10min à 12 000g. On utilise des cuves Hellma[®] de type 105.202-QS correspondant à un volume de 100 μl et un chemin optique de 10mm. Avant chaque mesure, la composition du tampon est entrée dans le logiciel d'acquisition et de traitement Dispersion Technology Software 4.10 (Malvern Instrument Ltd.) afin d'évaluer une valeur théorique de la viscosité de la solution. Une mesure dure le temps que 20 à 30 acquisitions reproductibles et cohérentes entre elles, soient accumulées. Les expériences de DLS ont été effectuées pour évaluer la qualité des échantillons de protéine purifiés et pour comparer, d'une purification à une autre, ces échantillons. Ces analyses sont faites notamment avant les tests de cristallisation et les mesures de dichroïsme circulaire.

2.2.3 CARACTERISATION DES INTERACTIONS PAR CALORIMETRIE A TITRAGE ISOTHERME

i. Préparation des peptides

Les peptides contenant un motif MIP sont synthétisés chez Genecust. Les peptides sont choisis avec le motif MIP, RSK(Y/F)F en leur centre et une longueur comprise entre 10 et 16 acides aminés. Le choix de la longueur du peptide est effectué selon la séquence présente de part et d'autre du motif MIP sur les différentes protéines. La règle suivie selon le fournisseur afin d'avoir un peptide soluble est que ce peptide ait moins de 50% de résidus hydrophobes et une charge totale nette différente de zéro. La charge nette est définie par :

$$S = -(D + E) + (R + K + H) \quad (6)$$

où D,E,R,K et H représentent respectivement le nombre d'acides aspartiques, d'acides glutamiques, d'arginines, de lysines et d'histidines dans la séquence du peptide.

Les peptides sont préparés avec une extrémité N-terminale acétylée et une extrémité C-terminale amidée selon les recommandations du fournisseur pour renforcer la stabilité de ces peptides vis-à-vis de protéases contaminantes. Ils sont resuspendus dans l'eau pour préparation injectable (ppi) à une concentration de départ de 10 ou 20mg/ml. La concentration des peptides est alors analysée sur

Nanodrop et par composition en acides aminés en utilisant le service d'analyses du SIMOPRO (iBitecs, CEA, Saclay).

Si les peptides ne sont pas totalement solubles, ils sont d'abord soniqués. Puis suivant leur score de solubilité S:

- si $S \geq 1$, on ajoute 10% d'acide acétique voire de TFA si ce n'est pas suffisant;
- si $S \leq 1$, on ajoute 10% d'hydroxyde d'ammonium.

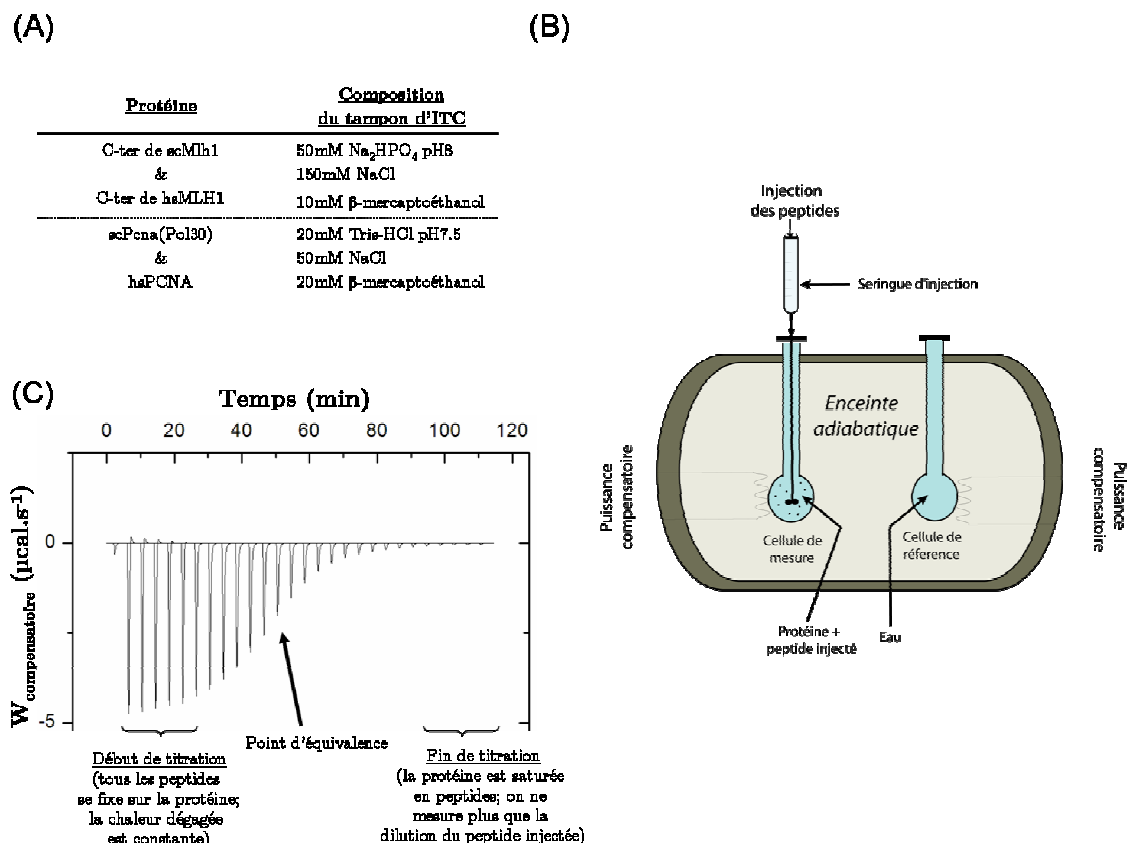
Si le peptide reste partiellement insoluble, on ajoute au maximum 5% de DMSO. (Source: 21st Century Biochemicals).

ii. Description du calorimètre et déroulement d'une mesure d'interaction

L'échantillon de protéine utilisé en calorimétrie est dialysé 3 fois pendant 2h dans le tampon correspondant (**Figure 2 - 4(A)**). La solution mère de peptide, concentrée environ 20 fois, est diluée dans le tampon de dialyse de la protéine afin d'avoir la protéine et le peptide dans des tampons aussi proches que possible. Les deux solutions (protéine et peptide) sont dégazées sous vide pendant 10min.

Le calorimètre utilisé est un système VP-ITC (MicroCal) (**Figure 2 - 4(B)**). Il mesure le dégagement ou l'absorption de chaleur, selon que la réaction étudiée est respectivement exothermique ou endothermique. La protéine est placée dans la cellule de mesure (de volume 1.4ml) sous agitation continue. Le peptide est injecté dans la cellule à intervalle régulier à l'aide d'une seringue de 300µL par une série d'injections de 10 ou 12µl espacées d'un temps suffisant (3min30 ou 4min). A chaque injection, l'interaction entre la protéine et le peptide, si elle a lieu, s'accompagne d'un dégagement ou d'une absorption de chaleur qui entraîne une variation de la température de la cellule de travail. Afin de rééquilibrer la température entre les deux cellules, l'appareil applique une puissance électrique W (en $\mu\text{cal.s}^{-1}$) qui est mesurée par le microcalorimètre.

Au début de la titration, la protéine est essentiellement sous sa forme libre. Tout le peptide injecté interagit avec la protéine et on observe des premiers pics d'échange de chaleur. Ces pics sont négatifs en cas de réaction exothermique (cas le plus fréquent) ou positifs en cas de réaction endothermique. Après un certain nombre d'injections, la moitié des molécules de protéines est déjà occupée par un peptide, de telle sorte que les pics observés sont diminués d'environ un facteur 2. En fin de titration, tous les sites de liaison sont déjà occupés et chaque nouvel ajout de peptide entraîne des dégagements de chaleur semblables à ceux observés avec le peptide seul contre le tampon. Ces pics correspondent aux dégagements de chaleur dus à la dilution du peptide dans la cellule de travail (**Figure 2 - 4(C)**).



iii. Détermination des paramètres thermodynamiques de l'interaction

L'énergie dégagée lors de l'interaction est enregistrée en fonction du temps puis intégrée par le logiciel Origin 7.0 (MicroCal). L'isotherme de titration qui s'ajuste sur la courbe expérimentale permet de déterminer la variation d'enthalpie de la liaison, ΔH° (kcal/mol), la constante d'association, K_a (M^{-1}) ainsi que la stoechiométrie d'interaction (n) (Figure 2 - 5). Les autres paramètres thermodynamiques, à savoir les variations d'entropie (ΔS°) et d'enthalpie libre (ΔG°) ainsi que la constante de dissociation (K_d) peuvent être déduits des relations:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = -RT \ln(K_a) = RT \ln(K_d) \quad (7)$$

avec T (K), la température absolue;

R , la constante des gaz parfaits, $1.98 \text{ cal.mol}^{-1}.\text{K}^{-1}$

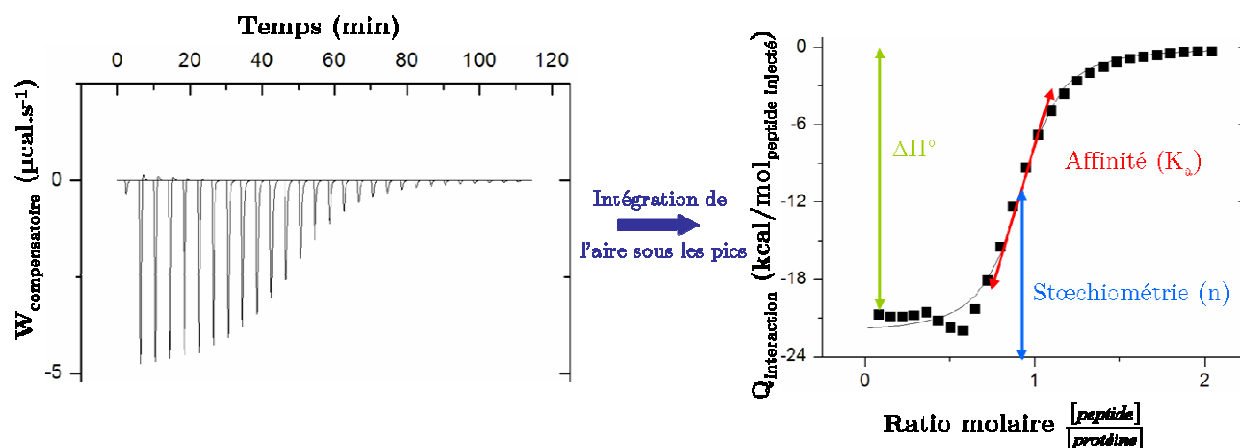


Figure 2 - 5: Interprétation d'un thermogramme entre la région C-terminale de hsMLH1 et le peptide 13mer pNtg2; détermination de la variation d'enthalpie, de l'affinité et de la stœchiométrie à partir de l'isotherme de titration

La force des interactions protéines-peptides étudiées est reflétée par la valeur ΔG° , l'interaction étant d'autant plus forte que cette valeur est négative. L'énergie ΔG° est la résultante de deux contributions énergétiques (Holdgate, 2001):

- Une contribution enthalpique ΔH° qui comprend notamment le nombre de liaisons hydrogène et d'interactions de Van der Waals au niveau de l'interface protéine-peptide étudiée;
- Une contribution entropique $-T\Delta S^\circ$ qui comprend notamment les interactions mettant en jeu des acides aminés hydrophobes, des changements conformationnels dus à l'interaction ainsi que par le relargage de molécules d'eau ou d'ions du fait de l'interaction.

Les principaux avantages de la microcalorimétrie isotherme de titration (ITC ou Isothermal Titration Calorimetry en anglais) sont l'absence de marquage des molécules, l'analyse de l'interaction en solution et non sur un support comme en résonance plasmonique de surface, l'accès à la stœchiométrie et aux paramètres thermodynamiques de l'interaction étudiée. L'inconvénient majeur de cette technique est sa faible sensibilité et la nécessité d'utiliser de grandes quantités de protéine et ligand (environ 1mg par courbe dans le cadre de ce projet).

2.3 ETUDE DES STRUCTURES TRIDIMENSIONNELLES

2.3.1 MODELISATION PAR HOMOLOGIE DU DOMAINE C-TERMINAL DU COMPLEXE MLH1*PMS1

i. Alignement de séquences

Afin d'optimiser le profil de sélectivité des séquences de MutL, à savoir inclure les homologues lointains sans inclure de faux positifs, nous avons utilisé la séquence en acide aminé de la région C-terminale de MutL (résidus 428 à 618) afin de générer

un profil contre la base de donnée non redondante (nr) du NCBI, en filtrant à une identité de séquence de 70%. Pour cela, nous avons utilisé PSI-BLAST (Altschul et al., 1997).

L'approche de PSI-BLAST a l'avantage d'être relativement rapide et simple puisqu'elle consiste à utiliser la séquence d'intérêt comme référence de l'alignement et à ajouter les séquences homologues une à une en intégrant les insertions telles qu'elles sont définies par l'alignement par pair. Cette stratégie génère un grand nombre d'insertion, ce qui rend les alignements multiples difficilement exploitable mais permet de détecter les homologues de MutL.

Les recherches d'homologues des protéines Mlh1 et Pms1 ont été réalisées avec le programme PSI-BLAST et plusieurs itérations (Altschul et al., 1997). Un sous-ensemble de 50 séquences de Mlh1 et 50 séquences de Pms1 eucaryotes ont été alignées avec le programme itératif MUSCLE (Edgar, 2004).

ii. Prédiction de structures secondaires et modélisation par homologie

Les alignements précédemment obtenus sont utilisés pour prédire les structures secondaires des régions C-terminales de scMlh1 et scPms1 à l'aide du serveur PSIPRED (McGuffin et al., 2000). Nous avons alors aligné les structures secondaires de Mlh1 et Pms1 prédites par PSIPRED avec les éléments de structures secondaires observés dans la structure cristallographique de MutL;

Cet alignement est ensuite utilisé pour modéliser les régions C-terminales de scMlh1 (de E528 à R768) et scPms1 (de F671 à K869) par homologie avec la structure cristallographique du domaine C-terminale de ecMutL et à l'aide du programme *Modeller* (Eswar et al., 2007).

2.3.2 MODELISATION DU PEPTIDE MIP DE EXO1 SUR LA SURFACE DE LA REGION C-TERMINALE DE MLH1

La procédure suivie pour modéliser l'ancrage d'un peptide contenant un motif MIP sur la structure cristallographique de la région C-terminale de Mlh1 a été défini par B. Gilquin au laboratoire: Nous avons en premier choisi de modéliser l'interaction avec le peptide d'Exo1 de 16 acides aminés dont l'interaction a été mesurée par ITC. Il s'agit du peptide ₄₃₉TLKDTRSKFFNKPSMT₄₅₄. La modélisation se fait par amarrage flexible (Bonvin, 2006), c'est-à-dire que la structure de la protéine est figée jusqu'à l'angle χ_1 des chaînes latérales. La procédure de modélisation moléculaire est une dynamique moléculaire à quatre dimensions (Crippen and Havel, 1990) sous contraintes ambiguës. La dynamique moléculaire à quatre dimensions est une modification de la fonction d'énergie implémentée dans CHARMM version c32b1 (Brooks et al., 2009) et du champ de force CHARMM27 afin d'ajouter une nouvelle dimension aux coordonnées cartésiennes (x,y,z) utilisées par l'algorithme de Verlet. L'ajout d'un degré de liberté réduit le nombre de minima locaux dans la fonction

d'énergie potentielle et donc, permet de franchir des barrières énergétiques présentes en 3 dimensions.

Les contraintes ambiguës imposées lors de l'ensemble de la dynamique moléculaire sont:

- les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E du peptide et les atomes C_γ , C_δ et C_ϵ des résidus F578^M, Y630^M ou Y631^M ou les atomes C_γ et C_δ de I689^M;
- les atomes C_γ , C_δ et C_ϵ de la phénylalanine F448^E du peptide et les atomes C_γ , C_δ et C_ϵ des résidus F578^M, Y630^M ou Y631^M ou les atomes C_γ et C_δ de I689^M;
- les atomes C_γ , C_δ et C_ϵ de la phénylalanine F448^E du peptide et les atomes C_γ , C_δ et C_ϵ des résidus F578^M, Y630^M ou Y631^M ou les atomes C_γ et C_δ de I689^M;
- les atomes C_γ , C_δ et C_ϵ du résidu Y631^M et les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E ou F448^E du peptide;
- les atomes C_γ , C_δ et C_ϵ du résidu Y631^M et les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E ou F448^E du peptide;
- les atomes C_γ , C_δ et C_ϵ du résidu F578^M et les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E ou F448^E du peptide;
- les atomes C_γ et C_δ du résidu I689^M et les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E ou F448^E du peptide;
- les atomes C_γ , S_δ et C_ϵ du résidu M626^M et les atomes C_γ , C_δ et C_ϵ de la phénylalanine F447^E ou F448^E du peptide.

Ces contraintes ont pour principal objectif de placer les deux résidus aromatiques du peptide à proximité des acides aminés hydrophobes du site S2. On ajoute 5 autres contraintes afin que les résidus aux extrémités du peptide épousent la forme du sillon de scMlh1:

- le squelette du peptide et la chaîne latérale du résidu I514^M (dans le but de faire un contact avec le résidu L440^E du peptide);
- le squelette du peptide et la chaîne latérale du résidu V521^M (d'un bord du sillon, à une extrémité);
- le squelette du peptide et la chaîne latérale du résidu L575^M (de l'autre bord du sillon, à la même extrémité);
- le squelette du peptide et la chaîne latérale du résidu I616^M (d'un bord du sillon, à l'autre extrémité);
- le squelette du peptide et la chaîne latérale du résidu W679^M (de l'autre bord du sillon, à l'autre extrémité).

Enfin, lors du refroidissement de la dynamique moléculaire, on ajoute une contrainte sur la chaîne latérale de la sérine du motif MIP et de tous les atomes d'oxygène du domaine C-terminal de scMlh1 afin d'orienter la chaîne latérale de la sérine vers le site S2.

La procédure consiste en une dynamique moléculaire à haute température (500K) en 14 000 étapes (sous les contraintes précédentes) puis en 40 000 étapes de refroidissement à 300K. Chaque étape dure 2fs. On génère ainsi 2 200 structures.

On sélectionne 50 structures qui ont une énergie de contraintes minimum (donc qui répondent le mieux aux contraintes d'un point de vue global).

2.3.3 LA CRISTALLOGRAPHIE AUX RAYONS X DES PROTEINES

i. Principe de la cristallographie: de la nécessité du cristal

La cristallographie aux rayons X est une technique physique permettant d'observer la structure des macromolécules biologiques. La technique s'apparente au principe de fonctionnement d'un microscope optique : lorsque la lumière traverse l'objet observé, les rayons lumineux sont diffusés, puis reconcentrés par une lentille, ce qui permet de reconstituer l'image. Le même principe peut s'appliquer à l'observation des atomes, mais dans ce cas, il faut utiliser des ondes électromagnétiques dont la longueur d'onde est de l'ordre de la distance qui sépare deux atomes liés covalentement. Pour visualiser l'arrangement des atomes dans une molécule, on utilise donc les rayons X dont la longueur d'onde est de l'ordre de l'Angström (10^{-10}m). Cependant, il n'existe pas à ce jour de matériau qui possède un indice de réfraction suffisamment différent de 1 pour être utilisé avec les rayons X durs. Donc, il n'est pas possible de refocaliser les rayons X et l'on peut simplement récupérer l'image obtenue au plan focal (le cliché de diffraction), perdant au passage l'information concernant la phase de l'onde diffractée. Ces phases doivent être obtenues suivant une autre approche et une fois ces phases obtenues, on applique un traitement mathématique basé sur la transformation de Fourier, pour obtenir l'image réelle. La transformée de Fourier de ces données génère l'image du nuage d'électrons qui correspond à la molécule cristallisée. Cependant, la diffusion des rayons X par une molécule isolée est trop faible pour fournir suffisamment de signal, ce qui oblige à faire diffuser un grand nombre de molécules simultanément afin d'additionner leurs contributions. Pour avoir une diffusion cohérente de toutes les molécules, celles-ci doivent être ordonnées dans l'espace. Cet état est atteint dans les cristaux qui sont constitués par un agencement périodique de molécules.

ii. De la diffraction à l'équation de la densité électronique

Le cristal peut être vu comme le produit de convolution d'un motif (une ou plusieurs macromolécules) par un réseau (caractérisé par la maille et le groupe d'espace du cristal). Le motif est répété à chaque nœud du réseau suivant les symétries du groupe d'espace. Lorsqu'un cristal est soumis aux rayons X, une partie du faisceau va être diffractée. La diffraction est due aux interférences constructives qui existent dans des directions particulières, en fonction de la maille du cristal et de

son orientation. Les conditions générales de la diffraction sont définies par la loi de Bragg (Bragg, 1913):

$$2d_{hkl} \sin \theta = n \cdot \lambda \quad (8)$$

où d_{hkl} , la distance séparant deux plans successifs du réseau cristallin d'indice $\{h,k,l\}$;

λ , la longueur d'onde du faisceau de rayons X;

θ , l'angle de réfraction du faisceau de rayon X;

n , un nombre entier.

Les directions des rayons X diffractés, appelées réflexions, dépendent uniquement du réseau cristallin et non de la structure du motif. Le réseau cristallin produit par diffraction un réseau appelé réseau réciproque. Les paramètres du réseau réciproque sont l'inverse du réseau cristallin. Les rayons X étant diffusés par les électrons, la diffusion de chaque atome est proportionnelle à son numéro atomique: plus l'atome aura d'électrons dans ses orbitales atomiques, plus fort sera son pouvoir de diffusion. Par conséquent, la structure déterminée par diffraction des rayons X est celle de la densité électronique des molécules dans l'unité asymétrique du cristal.

Chaque cliché de diffraction est composé de taches discrètes: les positions des taches c'est-à-dire leurs coordonnées $\{h,k,l\}$ dans le réseau réciproque, dérivent du réseau cristallin (paramètres de maille et groupe d'espace); l'intensité de l'ensemble des taches renseigne sur le contenu de la maille et donc sur la disposition des atomes dans la macromolécule. Ainsi, l'intensité I_{hkl} de chaque réflexion dans le spectre de diffraction est reliée à l'amplitude F_{hkl} de l'onde diffractée dans l'espace réciproque, appelé facteur de structure, par:

$$I_{hkl} = \vec{F}_{hkl} \cdot \vec{F}_{hkl}^* = F_{hkl}^2 \quad (9)$$

Le passage de l'ensemble des facteurs de structure à la densité électronique se fait selon une transformée de Fourier:

$$\begin{aligned} \rho_{xyz} &= \frac{1}{V} \sum_h \sum_k \sum_l \vec{F}_{hkl} \exp[-2i\pi(hx + ky + lz)] \\ &= \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} \exp[i\alpha_{hkl}] \exp[-2i\pi(hx + ky + lz)] \end{aligned} \quad (10)$$

En conséquence, pour déterminer la densité électronique de la protéine, il faut déterminer l'amplitude et la phase F_{hkl} et α_{hkl} des facteurs de structure de chaque $\{h,k,l\}$. Lors de la collecte des données de diffraction, le cristallographe récupère l'amplitude et perd l'information sur la phase α_{hkl} : c'est le "problème de la phase", discuté au §*vi*.

iii. Cristallogenèse

La cristallisation d'une macromolécule est le passage de cette dernière d'un état liquide vers un état solide, organisé grâce à des interactions non-covalentes. L'obtention des premiers cristaux d'une protéine est largement réalisée de façon

empirique et représente bien souvent une étape limitante dans la détermination de structure des protéines par cristallographie. La cristallisation d'une protéine passe par l'utilisation de conditions et d'agents plus ou moins éloignés des conditions physiologiques de la protéine: le pH, la concentration protéique, la température, la présence de solvant organique, de sels, de polymères organiques, etc. La mise au point de nouvelles conditions de cristallisation et l'amélioration de celles-ci est donc un mélange d'essais-erreurs et de contrôle rationnel de la croissance du cristal.

Plusieurs méthodes ont été mises au point afin de décrire le diagramme de solubilité protéine/eau/agent précipitant (**Figure 2 - 6**) (Asherie, 2004).

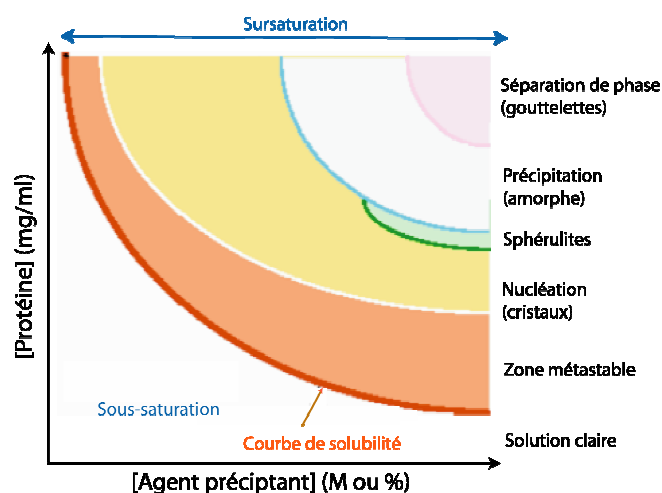


Figure 2 - 6: Schéma représentatif du diagramme de solubilité des protéines: état d'une protéine en fonction de la concentration en protéine et en agent précipitant

Le but des essais de cristallisation est d'atteindre lentement la zone de sursaturation afin que les protéines commencent à s'ordonner (et non à précipiter) et forment les premiers germes cristallins. Cette étape est appelée la nucléation. Pour atteindre cette zone, il faut d'une part que la concentration protéique soit suffisante et d'autre part ajouter des agents précipitants pour moduler le rapport entre les forces attractives et répulsives entre les macromolécules. Ces agents précipitants sont de trois natures:

- Les sels (sulfate d'ammonium, phosphate de sodium, ...) qui modifient la force ionique du milieu;
- Les polymères (PEG ou polyéthylène glycols) qui agissent sur la structuration de la couche de solvation de la protéine;
- Les solvants organiques (MPD ou Methylpentanediol, isopropanol, ...) qui agissent sur la constante diélectrique du milieu.

Différentes techniques de cristallisation permettent pour une concentration en protéine et un agent précipitant donnés, de parcourir le diagramme de solubilité différemment. Pour obtenir les cristaux de ce projet, deux méthodes ont été utilisées:

- La diffusion de vapeur en goutte assise qui a lieu entre une goutte contenant la protéine et un réservoir contenant la solution précipitante. La goutte est réalisée en

mélangeant 1.5µl de la solution de protéine et 1.5µl de la solution du réservoir. Au départ, la goutte est deux fois plus diluée que le réservoir en agent de cristallisation. L'expérience est fermée hermétiquement par une lamelle. Les composés volatiles de la goutte (essentiellement l'eau ou les agents organiques) vont s'évaporer jusqu'à équilibration de la pression de vapeur saturante. Au cours de cette évaporation, la concentration de la protéine et de l'agent précipitant augmentent progressivement (de A → B sur le schéma ci-dessous) et dans le cas favorable, des cristaux se forment et grossissent jusqu'à ce que la concentration en protéine atteigne la courbe de solubilité (de B → C) (Figure 2 - 7);

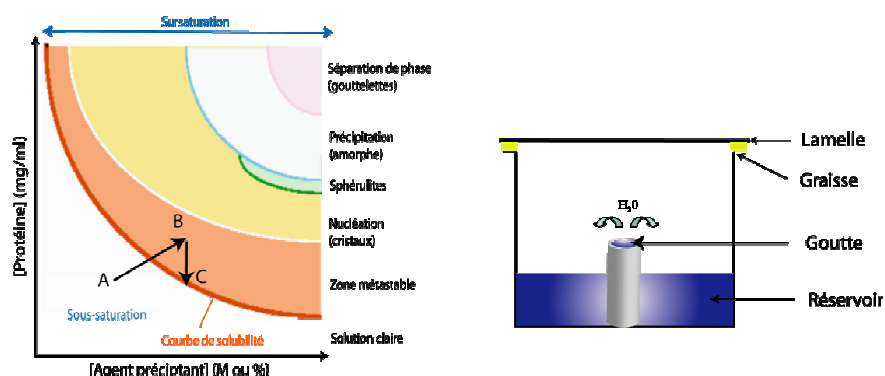


Figure 2 - 7: A droite, représentation du chemin d'équilibration au cours d'une expérience de cristallisation par diffusion de vapeur; à gauche, dispositif expérimental de cristallisation par diffusion de vapeur en goutte assise

- La cristallisation sous huile où sous un film d'huile de paraffine (Hampton Research), 1.5µl de solution de protéine est mélangé avec 1.5µl de solution précipitante de façon à atteindre directement la zone de sursaturation (en A) (Figure 2 - 8). Les cristaux vont grossir, comme pour la diffusion de vapeur, jusqu'à ce que la concentration en protéine atteigne la courbe de solubilité (de A → B). L'avantage est que la protéine n'est jamais en contact avec l'air évitant ainsi tout phénomène d'oxydation éventuel de la protéine; la croissance peut être très rapide car la protéine est mise en condition de sursaturation au moment où se forme la goutte.

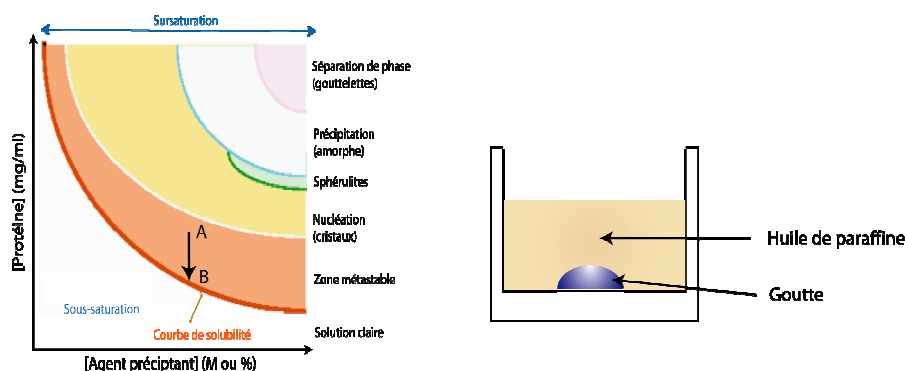


Figure 2 - 8: A droite, représentation du chemin d'équilibration au cours d'une expérience de cristallisation sous huile; à gauche, dispositif expérimental de cristallisation sous huile

Les cribles initiaux des conditions de cristallisation ont été effectués au cours de cette étude sur la plateforme HTX (High Throughput Crystallisation) dirigée par J. Marquès (PSB, PCUBE, Grenoble) (Dimasi et al., 2007). La plateforme dispose d'un robot de cristallisation PixSys 4200 (Genomic Solutions, U.K.) qui utilise des plaques Greiner Crystal Quick de 96 puits en gouttes assises. Chaque réservoir peut recevoir trois échantillons différents. Pour chaque échantillon, on teste au minimum 576 conditions de cristallisation (cribles Hampton Research et/ou Qiagen). Chaque goutte est composée de 100nl de solution de protéine et de 100nl de solution précipitante; les boîtes de cristallisation sont stockées dans un robot de visualisation de 500 boîtes à 4°C. Des images sont prises à intervalle régulier les quinze premiers jours et sont accessibles par internet depuis notre laboratoire.

Les résultats des cribles sont analysés afin d'identifier soit des macrocristaux, soit des pistes de cristallisation soit des précipités d'apparence cristalline ou des précipités suggérant l'agrégation de la protéine. Une seconde phase consiste alors à la reproduction des cristaux dans des gouttes de tailles supérieures et à l'optimisation de l'étape de croissance des cristaux afin d'obtenir des macrocristaux compatibles avec des analyses de diffractions au synchrotron.

iv. Optimisation par ensemencement des cristaux

Les premiers cristaux obtenus pour le complexe Mlh1*Pms1 sont de fines aiguilles qui poussent en grand nombre et très rapidement. Afin d'optimiser leur taille et potentiellement leur pouvoir de diffraction, nous avons utilisé la technique de microensemencement.

En effet, la nucléation nécessite une forte sursaturation alors que la croissance peut continuer à des concentrations plus faibles en agent précipitant: sur le diagramme de solubilité, il existe une zone dite métastable où en fonction de la cinétique de l'expérience, les cristaux n'apparaissent pas toujours spontanément mais où ils peuvent grossir. L'ensemencement consiste à récolter des cristaux dans des gouttes où ils sont apparus mais où ils n'ont pas poussé de façon satisfaisante et à les déposer dans des gouttes qui, à l'équilibre, sont dans cette zone métastable. Le microensemencement permet donc d'optimiser séparément la nucléation et la croissance.

Dans un premier temps, les microcristaux sont récupérés dans la goutte et dilués dans 50µl de la solution de cristallisation présente dans le réservoir de la goutte. Les cristaux sont centrifugés avec une bille à 10 000 g pendant 10min puis broyés au vortex entre 20s à 2min. On ajoute alors de 0.3µl (pour les premiers essais) à 0.1µl (pour l'optimisation de la croissance) de cette solution à une goutte préalablement équilibrée et ayant une concentration en agent précipitant légèrement inférieure à celle utilisée pour les gouttes où la protéine commence à cristalliser spontanément. L'ajout de quelques germes permet dans certains cas d'avoir des cristaux de taille supérieure grâce à une croissance plus lente et plus régulière.

Nous avons observé au cours de ces expériences que les cristaux poussent fréquemment à la surface des microcristaux ensemencés mais également à l'interface solution précipitante/air/plastique. Nous avons alors choisi d'essayer une approche récemment proposée d'introduction d'agents de nucléation hétérogènes (Thakur et al., 2007). Nous avons essayé au laboratoire trois des agents suggérés (billes d'hydroxyapatite (Sigma-Aldrich® -ref 677418), crins de chevaux broyés, algues séchées et broyées). Dans les trois cas, une pointe de spatule d'agent de nucléation hétérogène est ajoutée à 100µl de solution de cristallisation. L'ensemble est agité au vortex. On ajoute de 0.1 à 0.3µl à la goutte de cristallisation préalablement équilibrée. Cette technique a permis notamment dans le cas de l'hydroxyapatite, de mieux contrôler le nombre de germes de nucléation et leur croissance.

v. Cryocongélation et collection des données

L'analyse des premières aiguilles du complexe entre les régions C-terminales de Mlh1 et Pms1 sur les lignes des synchrotrons Soleil (Gif-sur-Yvette) et de l'ESRF (Grenoble) a montré que ces cristaux ont un faible pouvoir diffractant (2.6Å au mieux). La faible diffraction de cristaux peut dans certains cas provenir d'une mauvaise congélation. En effet, le rayonnement synchrotron étant particulièrement brillant, il est nécessaire de collecter les données à températures cryogéniques (Garman and Owen, 2006). Si ce n'est pas le cas, les rayons endommagent fortement et rapidement le cristal en créant des radicaux libres. Ces radicaux vont alors migrer au sein du cristal vers les sites de haute affinité électronique (ponts disulfures, métaux, ...). Le principal problème lorsque l'on congèle les cristaux de protéines est la formation de glace qui endommage le réseau cristallin et provoque l'apparition d'anneaux sur les clichés de diffraction (notamment des anneaux à 3.5, 3.7, 3.9Å). De nombreuses solutions cryoprotectantes ont été essayées: polyéthylène glycol à petit poids moléculaires, alcools (glycérol, éthylène glycol) et sucres (xylitol, glucose) ainsi que le LiCl à haute concentration. Au final, pour les différents cristaux obtenus, les cryoprotectants les plus efficaces ont été le glycérol et le polyéthylène glycol additionnés à hauteur de 20-30% (v/v) à la solution de cristallisation.

Les données ont été collectées sur les lignes de lumière Proxima1 (Synchrotron Soleil), ID23-1&2 et ID29 (ESRF). Du fait du faible pouvoir diffractant et de la très grande anisotropie de diffraction (souvent un facteur 2 en résolution entre la meilleure et la pire orientation), plus de 330 cristaux (toute forme confondue) ont été testés.

Les jeux de données natifs ont été collectés à une longueur de 0.98Å (12.3keV) avec un détecteur Quantum 315r (ADSC). Les cristaux refroidis grâce à un flux d'azote gazeux à 100K, ont été exposés entre 0.5 et 1s au faisceau de rayons X. Chaque cliché de diffraction correspond à une rotation de 0.5 ou 1°. Le nombre d'images effectuées est fonction de la durée de vie du cristal et du groupe d'espace.

vi. *Traitement des données*

Les clichés de diffractions sont analysés avec le programme XDS (Kabsch, 1988) qui présente l'avantage de calculer un profil tridimensionnel des taches de diffraction, permettant une meilleure intégration de celles-ci.

L'intégration se déroule en trois étapes:

- L'indexation (IDXREF): à partir des réflexions les plus fortes (COLSPOT), le programme calcule les vecteurs entre les taches de diffraction et détermine le réseau de Bravais et la maille;
- L'intégration (INTEGRATE): on mesure l'intensité de chaque tâche,;
- La mise à l'échelle (CORRECT ou *Scala* (Evans, 2006)): le programme corrige diverses erreurs expérimentales en se basant sur les intensités mesurées pour des réflexions équivalentes par symétrie ou enregistrées plusieurs fois. Cette étape corrige par exemple l'intensité moyenne du rayonnement incident qui peut varier pour chacune des images de l'enregistrement.

vii. *Résolution du problème de la phase*

Comme mentionné précédemment, une étape cruciale de la cristallographie aux rayons X (avec celle de l'obtention des cristaux) est l'obtention des phases de chaque réflexion $\{h,k,l\}$. Contrairement aux petites molécules organiques, les méthodes directes ne sont, en général, pas utilisables pour les protéines. Il existe alors deux méthodes principales pour surmonter le problème des phases selon que l'on dispose d'une structure homologue (méthode du remplacement moléculaire) ou que l'on peut fixer des atomes «lourds» dans le cristal (méthode des dérivés lourds et des sélénométhionines notamment).

- **Le remplacement moléculaire**

Lorsque l'on dispose d'un modèle (comme la structure cristallographique du domaine C-terminal de ecMutL dans notre cas), le remplacement moléculaire (Rossmann, 1990) permet de positionner un ou plusieurs modèles dans la maille du cristal étudié. Le positionnement de ces modèles permet alors de calculer une première estimation des phases pour les facteurs de structure du cristal étudié. Pour un gain de temps et d'efficacité, la plupart des programmes actuels dont *MolRep* (Vagin and Teplyakov, 2009), décomposent le positionnement de la molécule modèle dans le cristal en deux étapes : la recherche des angles de rotation et des vecteurs de translation qui permettent de passer de la position initiale du modèle à la position de la molécule étudiée dans le cristal. Les deux étapes consistent en une comparaison entre la fonction de Patterson calculée P_{calc} et celle observée P_{obs} . La fonction de Patterson est définie par (Blundell and Johnson, 1976):

$$\begin{aligned}
P(uvw) &= \int_V \rho(xyz) \cdot \rho(x+u, y+v, z+w) dV \\
&= \frac{2}{V} \sum_h \sum_k \sum_l F_{hkl}^2 \cdot \cos[2\pi(hu + kv + lw)]
\end{aligned} \tag{11}$$

avec V, le volume de l'unité asymétrique. Elle représente donc la carte des vecteurs interatomiques calculables à partir des seules intensités. On déduit alors l'expression des fonctions de rotation (RF) et de translation (TF):

$$\begin{aligned}
RF(R) &= \int_V P_{obs}(u) \cdot P_{calc}(u, R) dV \\
TF(t) &= \int_V P_{obs}(u) \cdot P_{calc}(u, t) dV
\end{aligned} \tag{12}$$

où R et t sont respectivement les matrices de rotation et de translation. Pour la fonction de rotation, le volume d'intégration est choisi de telle façon que le maximum de vecteurs intramoléculaires soit inclus tout en excluant un maximum de vecteurs intermoléculaires. Le rayon d'intégration de la fonction de rotation est donc généralement choisi proche du rayon de giration du modèle, supposant ainsi que la forme globale de la protéine étudiée est globulaire et de rayon proche de celui du modèle.

Outre le choix de ce rayon d'intégration, le modèle utilisé doit être aussi proche que possible de la structure de la protéine d'intérêt. On estime souvent qu'au-delà de 30% d'identité de séquence, le modèle et la protéine étudiée ont de grandes chances d'avoir un repliement similaire. Il est également possible de ne pas utiliser la résolution au-delà de 3Å du jeu de données natif afin de rendre plus "flou" les différences structurales entre le modèle de départ et la structure réelle. Pour un modèle de départ à faible homologie (ce qui est notre cas), le remplacement moléculaire peut être fait en utilisant les données de diffraction jusqu'à 6Å.

Dans les cas favorables, les solutions obtenues permettent d'obtenir des phases initiales et des cartes de densité électronique de qualités suffisantes pour reconstruire la protéine étudiée.

- Le phasage expérimental

Cette méthode tire parti des propriétés spécifiques de diffraction du cristal lorsque celui-ci contient des atomes dits «lourds» car contenant beaucoup d'électrons par rapport aux atomes des protéines (N, C, O, H et S).

Les premières méthodes proposées dans ce sens sont les méthodes de remplacement isomorphe (MIR et SIR) (Green et al., 1954) qui exploitent la contribution des atomes lourds dans les clichés de diffraction. Il est aussi possible de tirer parti d'un phénomène physique résultant de la diffusion d'un rayonnement X par des atomes : la diffusion anomal. Aujourd'hui, il existe des techniques exploitant les deux phénomènes, isomorphe et anormale (SIRAS ou MIRAS).

Pour résoudre la structure du complexe C-ter de scMlh1*scPms1, nous avons utilisé la méthode de diffusion anormale à une longueur d'onde (SAD) qui est détaillée dans le paragraphe suivant.

viii. La méthode de diffusion anormale à une longueur d'onde

- La diffusion anormale

Loin des conditions d'absorption d'un atome, le rayonnement incident ne touche que les électrons des couches supérieures considérés comme libre: ils réémettent une onde de même fréquence et de même amplitude que l'onde incidente. Mais à une longueur d'onde proche de celle d'absorption d'un atome lourd, les électrons des couches internes entrent en résonance, ce qui introduit un déphasage de l'onde émise et modifie l'amplitude de facteur de diffusion f (Cromer, 1965):

$$f = f_0 + f'_\lambda + if''_\lambda \quad (13)$$

où f_0 est le facteur de diffusion à une fréquence supérieure à celle d'absorption

f' et f'' sont les composantes réelle et imaginaire de la diffusion anormale et dépendent de la longueur d'onde λ .

Cette diffusion anormale étant relativement faible, il est nécessaire de connaître ses composantes f' et f'' avec précision. Or la composante imaginaire f'' est reliée au coefficient d'absorption μ par la théorie de Mie:

$$f''_\lambda = \frac{1}{2r_0\lambda} \frac{M}{N_A} \frac{\mu}{\rho} \quad (14)$$

avec r_0 , le rayon de l'électron;

λ , la longueur d'onde;

N_A , le nombre d'Avogadro;

M , la masse atomique de l'élément;

$\frac{\mu}{\rho}$, le coefficient d'atténuation massique.

Ce coefficient d'absorption peut être mesuré par un spectre de fluorescence qui correspond au passage des couches électroniques à haute énergie vers les couches internes (la ligne Proxima 1 est équipée d'un détecteur au silicium à diffusion).

La composante réelle du facteur de diffusion peut alors être déduite à partir de la relation de dispersion de Kramers-Kronig (James, 1958):

$$f'(E) = \frac{2}{\pi} \int f''(E_0) \frac{E_0^2}{E^2 - E_0^2} dE_0 \text{ et } E = \frac{hc}{\lambda} \quad (15)$$

avec E_0 et E , les énergies respectives de l'onde incidente et diffusée;

h , la constante de Planck;

c , la célérité de l'onde.

- Théorie de la détermination des phases expérimentales par la méthode SAD

Le facteur de diffusion d'un atome j de la protéine est relié au facteur de structure de cette protéine par:

$$F_{hkl} = \sum_j f_j \cdot \exp[2i\pi(hx_j + ky_j + lz_j)] \quad (16)$$

Or si on considère deux réflexions $\{hkl\}$ et $\{\bar{h}\bar{k}\bar{l}\}$, alors d'après la loi de Friedel,

$$|F_{hkl}^+| = |F_{\bar{h}\bar{k}\bar{l}}^-| \quad (17)$$

à condition que le facteur de diffusion f soit réel (donc que la protéine ne contienne pas de diffuseur anomal) (**Figure 2 - 9(A)**). Dans le cas où il y a de la diffusion anormale, la loi de Friedel n'est plus vraie (**Figure 2 - 9(B)**). Le facteur de diffusion présente une composante imaginaire.

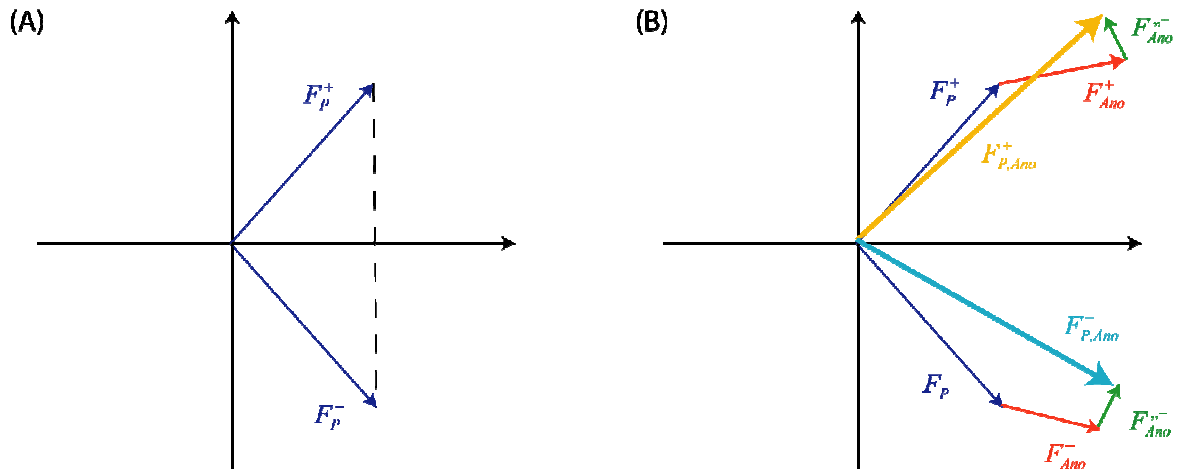


Figure 2 - 9: Représentation vectorielle des facteurs de structure (A) En l'absence de diffuseur anomal $|F_p^+| = |F_p^-|$ (B) En présence d'un diffuseur anomal qui induit une partie imaginaire F_{Ano}'' , ce qui implique $|F_{P,Ano}^+| \neq |F_{P,Ano}^-|$ (d'après (Taylor, 2003))

- Détermination des sites d'atomes lourds

Dans un premier temps, il nous faut déterminer les facteurs de structure des atomes lourds et ceci nécessite de connaître leurs positions. Dans notre cas, nous avons utilisé SHELXD (Dall'Antonia and Schneider, 2006) qui utilise l'approche *Shake-and-Bake*, une combinaison des propriétés de la fonction de Patterson et des méthodes directes.

La méthode directe peut être utilisée pour déterminer les sites d'atomes lourds car d'une part, les atomes lourds présents sont peu nombreux (<100 atomes dans le cas général) et d'autre part, ces atomes lourds sont en général relativement éloignés les uns des autres (>3Å). A partir d'un jeu de phases initiales, cette approche permet de compléter les phases d'atomes lourds en appliquant la relation de probabilité des phases à savoir la formule de la tangente (Karle and Hauptman, 1956):

$$\tan(\varphi_h) = \frac{\sum_k |E_k E_{h-k}| \cos(\varphi_k + \varphi_{h-k})}{\sum_k |E_k E_{h-k}| \sin(\varphi_k + \varphi_{h-k})} \quad (18)$$

où φ représente la phase et E est le facteur de structure normalisé. Elle permet donc de calculer la phase φ_h à partir d'une ou plusieurs paires de phases φ_k .

Dans SHELXD, le petit nombre de phases initiales est déterminé à l'aide de la fonction de Patterson: à partir d'un jeu de position d'atomes lourds, on calcule les vecteurs interatomiques que l'on compare à la carte de Patterson expérimentale.

Une fois la position des atomes localisée, on peut déterminer les phases de la molécule cristallisée (**Figure 2 - 10**). Le module du facteur de structure F_P de la protéine est connu par la mesure des intensités de diffraction des données natives. On peut tracer le cercle de rayon correspondant au module de F_P (en bleu). La position des atomes lourds donne accès à la fois au module et à la phase du facteur de structure F_{Ano} (vecteur vert). On connaît également le module du facteur de structure de la protéine avec le diffuseur anomal (en jaune). L'intersection entre les deux cercles bleu et jaune donne alors deux solutions possibles pour la phase de la protéine.

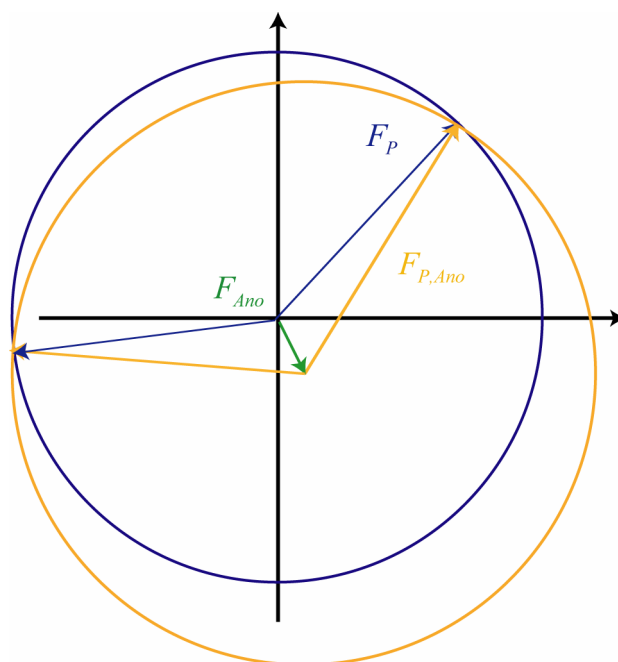


Figure 2 - 10: Construction de Harker montrant la détermination graphique de la phase grâce à la présence d'atomes lourds

- Levée de l'ambiguïté et amélioration des phases

Afin de lever l'ambiguïté entre les deux solutions, on peut utiliser une autre longueur d'onde (la méthode MAD). Dans ce projet les petits cristaux obtenus ont une durée de vie faible dans le faisceau synchrotron et notre faible symétrie oblige à collecter un nombre important d'images. De plus, la maille de nos cristaux suggérait

la présence d'un fort pourcentage de solvant, *i.e.* 69% selon l'analyse du coefficient de Matthews. Ces deux éléments nous ont fait nous orienter vers l'approche SAD qui nécessite un seul jeu de données redondant avec des cristaux de protéine marquée avec des sélénométhionines.

En théorie, le phasage avec un jeu de données SAD est possible car à la différence du remplacement isomorphe, les deux solutions du diagramme de Harker ne sont pas équiprobables (Hendrickson et al., 1990). En effet, l'atome diffuseur faisant partie de la protéine, il est probable que la phase de la protéine soit proche de la phase de l'atome lourds (Dodson, 2003). Le programme *Phaser* (McCoy et al., 2007) propose une approche statistique utilisant une fonction de distribution de probabilité spécifique pour la méthode SAD.

En pratique, cet effet est mineur et ne permet pas, en général, de lever l'ambiguïté. Les phases calculées par *Phaser* correspondent au centroïde des deux solutions possibles. La levée de l'ambiguïté intervient lors de la modification de la densité électronique. Pour ce projet, nous avons utilisé le programme *Parrot* (Cowtan, 2010) qui utilise les méthodes de modification de densité classiques (et non statistiques) c'est-à-dire:

- Le nivellement de solvant: les zones de solvant sont caractérisées par une densité électronique plus faible et plus constante. En localisant ces zones et en assignant la valeur moyenne de la densité à ces endroits, on peut améliorer la densité dans les zones de protéines;
- L'ajustement d'histogrammes de densité: les atomes et liaisons étant de nature globalement similaire pour toutes les protéines, on peut calculer un histogramme moyen de la répartition de la densité électronique dans les zones de protéines à partir d'une librairie de protéines. Cet histogramme a une répartition différente par rapport à une distribution de phases aléatoires. On peut donc appliquer une mise à l'échelle des phases expérimentales afin que la répartition de ces dernières ressemble à celle de phases bien définies.

Les phases calculées à l'aide de ces méthodes correspondent alors à une solution plus proche de la réalité que les phases expérimentales calculées par *Phaser*. Pour éviter les biais, ces phases sont combinées aux phases initiales avec un poids correspondant à la figure de mérite. Au final, en effectuant plusieurs cycles de modifications de densité électronique, l'une des deux solutions va ressortir au détriment de l'autre et ceci nous permet de lever l'ambiguïté.

- Construction automatique dans les phases expérimentales

A partir de ces phases expérimentales améliorées, on peut commencer à construire un modèle. Comme le jeu de données des cristaux de la protéine marquée aux sélénométhionines, SeMet, est à 3.12Å de résolution, le programme le plus approprié pour construire automatiquement la protéine dans les cartes de densité électronique est *Buccaneer* (Cowtan, 2008). Ce programme représente un carbone C $_{\alpha}$

par sa densité électronique comprise dans une sphère de 4Å. *Buccaneer* cherche un point de départ qui correspond le mieux à cette sphère. Puis il essaie d'étendre cette chaîne dans chaque direction en prenant en compte la densité et les angles de Ramachandran (Phi, Psi). Enfin, il modélise les chaînes latérales par la densité électronique comprise dans une sphère de 5.5Å centrée sur le carbone C β , et ce, pour les 20 acides aminés (le rayon de 5.5Å correspondant à l'insertion de tous les atomes hors hydrogène, d'une chaîne latérale d'un tryptophane). De la même manière, *Buccaneer* essaie de placer ces chaînes latérales à partir du squelette construit précédemment.

Dans notre cas, le programme effectue trois cycles de ce type pour dix cycles d'affinement avec *Refmac* version 5.5 (Murshudov et al., 1997), et ce, 100 fois. Sur ces 100 cycles, on enregistre cinq structures correspondant à des moments où *Buccaneer* construit les plus longues chaînes. La structure la plus aboutie est améliorée dans les phases expérimentales à travers *Coot* version 0.6.1 (Emsley and Cowtan, 2004).

- Affinement et validation de la structure

La structure obtenue à partir des phases expérimentales est ensuite affinée d'abord contre les données anormales à 3,12Å puis dans les données natives à 2.60Å. Le premier cycle d'affinement est effectué à l'aide de *Phenix.refine* version 1.6 (Afonine et al., 2005). Un recuit simulé à 10 000K est utilisé afin de passer certaines barrières énergétiques. Nous avons alors effectué 93 cycles alternant affinement avec *Buster* version 2.8 & 2.9 (Bricogne et al., 2009) et reconstruction dans *Coot*. Le programme *Buster* utilise le maximum de vraisemblance et présente ainsi l'avantage de calculer des cartes moins biaisées que celles utilisant *Refmac* ou *Phenix.refine*. Chaque étape d'affinement avec *Buster* comporte 10 cycles. La fonction de géométrie impose des contraintes d'angles et de liaisons importées de la librairie EH99 (Engh and Huber, 2006) auxquelles on a ajouté des contraintes sur les torsions impropres autour des centres chiraux (afin de minimiser la distorsion des angles C β). Un second avantage du programme *Buster* est la qualité de l'affinement de la géométrie de la protéine.

La structure est ensuite affinée suivant les données SeMet pendant 100 cycles. Pour définir le désordre vibrationnel associé aux domaines de la protéine dans le cristal, on utilise la macro "TLSalternate" en définissant 8 domaines; ces 8 domaines ont été déterminés par analyse des facteurs de température B et à l'aide du serveur TLSMD (Painter and Merritt, 2006). Pour s'assurer de la stabilité de l'affinement, les coefficients des tenseurs thermiques $\vec{T}, \vec{L}, \vec{S}$ sont réinitialisés à chaque fois.

L'affinement du modèle est suivi et validé par les facteurs R et R_{free}. Ces deux facteurs permettent de contrôler l'accord entre les facteurs de structures observés (F_{obs}) et les facteurs de structures calculés à partir du modèle construit (F_{calc}) d'après la formule :

$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|} \quad (19)$$

Le facteur R diminue nécessairement au cours de l'affinement car on optimise l'accord entre les facteurs de structure observés et calculés. Le facteur R_{free} moins biaisé que le facteur R, est calculé sur un ensemble de réflexions exclues de la procédure d'affinement. Ces réflexions sont choisies de manière aléatoire et représentent 5 % des réflexions totales. Pour être sûr de ne pas biaiser ce facteur, il est importé et étendu à 2.60Å lors du passage des données SeMet aux données natives.

Les facteurs R et R_{free} finaux ainsi que les déviations moyennes des distances et les angles entre les liaisons sont calculés avec le programme *model_vs_data* (Afonine et al., 2010) qui tient compte des coefficients des tenseurs $\vec{T}, \vec{L}, \vec{S}$. La qualité géométrique du modèle final est déterminée par le serveur *MolProbity* (Chen et al., 2010).

3. ANALYSE FONCTIONNELLE ET STRUCTURALE DU DOMAINE C-TERMINAL DE MUTL α

3.1 IDENTIFICATION ET CARACTERISATION D'UN NOUVEAU SITE D'INTERACTION DE MLH1

3.1.1 IDENTIFICATION DU SITE D'INTERACTION DES PROTEINES CONTENANT UN MOTIF MIP PAR MUTAGENÈSE DIRIGÉE ET MODELISATION PAR HOMOLOGIE

Comme évoqué en introduction, le domaine C-terminal de Mlh1 est au cœur d'un réseau d'interactions avec au moins 7 protéines impliquées dans les mécanismes "3R" (Réparation, Réplication et Recombinaison). Le laboratoire de S. Boiteux² a ainsi découvert que chez *S. cerevisiae*, la glycosylase/lyase Ntg2 du BER interagit avec la région C-terminale de Mlh1 *via* une séquence appelée motif MIP et que ce motif est conservé chez l'exonucléase Exo1 et l'hélicase Sgs1. Le motif MIP est une séquence de cinq acides aminés, (R/K)SK(Y/F)F. Il est situé, dans les protéines Ntg2, Exo1 et Sgs1, au sein de régions prédites comme non structurées. On le retrouve ainsi en position N-terminale (23 à 27) du domaine glycosylase de Ntg2, en position C-terminale (444 à 448) du domaine exonucléase d'Exo1 et enfin en position C-terminale (1382 à 1386) du domaine hélicase de Sgs1. Ce motif est plus ou moins conservé au sein des séquences eucaryotes des protéines Ntg2, Exo1 et Sgs1. Nous avons retrouvé un motif MIP dans la région C-terminale d'Exo1 de nombreux organismes eucaryotes, de la levure à l'homme. A l'inverse le motif MIP de Ntg2 n'est retrouvé que chez quelques levures hemiascomycètes proches de *S. cerevisiae*. Enfin, la protéine Sgs1 est un cas intermédiaire car un motif MIP n'est retrouvé que chez un certain nombre de levures et d'animaux incluant l'hélicase humaine BLM.

Délétions des extrémités N- et C-terminales de la région (483-769) de Mlh1 de levure: Afin de mieux comprendre le rôle de ces interactions et leur possible compétition entre elles ou vis-à-vis d'autres partenaires, le premier objectif de ce projet a été de caractériser le site d'interaction des protéines contenant un motif MIP sur la région C-terminale de Mlh1. Pour ce faire nous avons choisi de travailler dans un premier temps sur le modèle de levure et d'utiliser une approche combinant mutagenèse dirigée et tests de double hybride au CEA de Fontenay-aux-Roses et modélisation par homologie au LBSR. L'étude initiale de l'interaction par Gellon et al. avait permis de délimiter le domaine de Mlh1 (483-769) comme un domaine suffisant pour l'interaction avec Ntg2, Exo1 et Sgs1 dans un test double hybride (Gellon et al., 2002). Plusieurs délétions des extrémités N- et C-terminales de la région (483-769) de

² Laboratoire de Radiobiologie de l'ADN, iRCM, CEA Fontenay-aux-Roses

Mlh1 ont alors été testées par double-hybride (**Figure 3 - 1**). Ces expériences montrent que seules les 10 derniers acides aminés (construction Mlh1 Δ C10) ne sont pas nécessaires pour l'interaction avec Ntg2, Exo1 mais également pour le site d'hétérodimérisation avec Pms1. Toutes les délétions testées à l'extrémité N-terminale annule l'interaction avec Exo1 et Ntg2. Deux de ces délétions (construction Mlh1 Δ N35 et Mlh1 Δ N45) maintiennent cependant l'interaction avec Pms1. Ce dernier résultat apporte plusieurs informations utiles. (i) Le maintien d'une interaction avec Pms1 pour les constructions Mlh1 Δ N35 et Mlh1 Δ N45 suggère que ces constructions de Mlh1 sont repliées. (ii) Les résultats observés avec ces délétions indiquent que le site d'hétérodimérisation, appelé site S1 par la suite, ne se superpose pas entièrement au site d'interaction avec les protéines contenant un motif MIP, appelé site S2 par la suite. (iii) Enfin ce résultat suggère que la région (483-518) contient vraisemblablement des acides aminés du site 2.

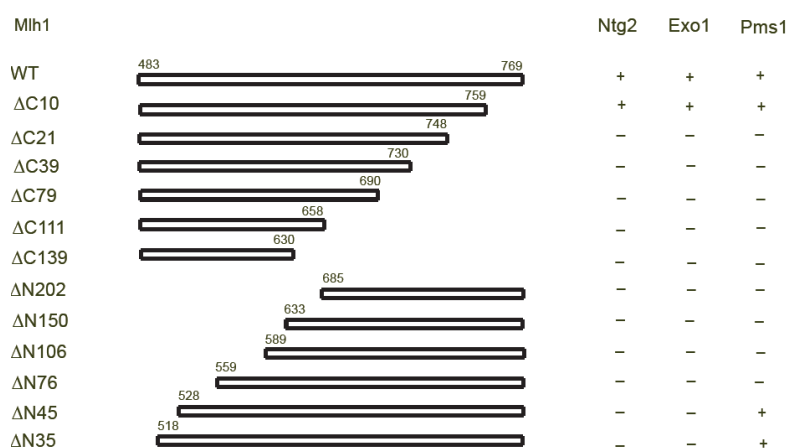


Figure 3 - 1 : Schéma des constructions du domaine C-terminal de Mlh1 testées en double hybride ; les essais de double-hybride sont effectués en utilisant Ntg2(1-380), Exo1(400-702) ou Pms1(661-873) comme appât. Les résultats sont indiqués comme (+) pour des colonies bleues (activité β -galactosidase en présence de X-gal) et (-) pour des colonies blanches après 24 heures d'incubation à 30°C

1ère série de mutants ponctuels sur des positions conservées de la région (483-769) de Mlh1 de levure: Etant donnée la forte conservation du motif MIP parmi les séquences eucaryotes d'Exo1, nous avons fait l'hypothèse que le site d'interaction de Mlh1 avec Exo1 et les autres protéines contenant un motif MIP met en jeu des acides aminés conservés de la région (483-769) de Mlh1. Nous avons également fait l'hypothèse que certains de ces acides aminés appartiennent à l'extrémité N-terminale (483-518) étant donné les résultats des tests de double-hybride présentés dans le paragraphe précédent. Vingt résidus conservés de la région C-terminale (CTD) (483-769) de Mlh1 ont été sélectionnés et mutés principalement en alanine. Ces mutants ont été testés en double-hybride pour l'hétérodimérisation et pour l'interaction avec Ntg2 et/ou Exo1 (**Tableau 3 - 1**). 17 mutants sur 20 conservent la capacité d'interagir à la fois avec Pms1 et avec Ntg2. 3 mutants, S513A, Y630A et E682A, interagissent avec Pms1 (démontrant que le domaine ainsi défini est correctement replié) mais

perdent la possibilité d'interagir avec Ntg2. Nous avons inclus dans cette série de mutants, un mutant R547A qui correspond à une position étudiée précédemment dans deux études sur Mlh1. La position R547 fait partie d'une séquence de 5 acides aminés, ₅₄₃DEERR, qui lorsqu'elle est mutée en alanine, annule l'interaction avec Pms1 (Argueso et al., 2003). Cette position correspond également à un simple mutant R547K identifié par le laboratoire de R. Kolodner comme mutant hypomorphe du MMR (Amin et al., 2001). La mutation R547A dans notre étude annule l'interaction avec Pms1 sans affecter l'interaction avec Ntg2. Pour un certain nombre de ces mutants, l'interaction a également été testée entre Mlh1 et Exo1.

Ces premiers résultats confirment l'existence de sites distincts d'interaction avec Pms1 (site S1, position R547) ou avec Ntg2 et Exo1 (site S2, positions S513, Y630 et Y682). Ils suggèrent également que le site S2 occupe au moins 3 zones distinctes de la région C-terminale de Mlh1 (autour des positions 513, 630 et 682).

Tableau 3 - 1: Résultats du 1^{er} tour de mutagenèse sur le domaine C-terminal de Mlh1 et des expériences de double-hybride pour l'interaction avec Pms1, Ntg2 et pour certains mutants avec Exo1; les interactions annotées (+) traduisent des colonies bleues, celles (+/-) traduisent des colonies bleues clair; celles (-) traduisent des colonies blanches; celles (n. d.) n'ont pas été déterminées

Mutation ponctuelle du CTD de Mlh1	Interaction en double hybride		
	Pms1	Exo1	Ntg2
S513A	+	-	-
Y538A	+	+	+
E544A	+	+	+
R547A	-	+	+
Q552A	+	+	+
F558A	+	+	+
C566S	+	+	+
Y571A	+	n. d.	+
Q572A	+	n. d.	+
I573E	+	+	+
Y630A	+	-	-
L655E	+	+	+
E682A	+	-	-
C685A	+	+	+
H733A	+	+	+
C738S	+	+	+
K740A	+	+	+
P746A	+	+	+
K751	+	+	+
E755A	+	+	+
C769S	+	+	+

Modélisation par homologie et proposition de nouveaux mutants : Afin d'identifier un plus grand nombre de résidus du site S2, nous avons choisi de réaliser un modèle par homologie de la région C-terminale de Mlh1 et d'analyser les acides aminés conservés susceptibles d'être situés à proximité des premiers acides aminés identifiés (S513, Y630 et E682). Nous avons fait l'hypothèse d'une homologie structurale globale de la région C-terminale de Mlh1 avec la région C-terminale de MutL d'*E. coli* (ou ecMutL), seule structure connue de CTD de MutL à ce moment de l'étude (Guarne et al., 2004). Le domaine tridimensionnel du CTD de Mlh1 a été construit à l'aide du programme MODELLER et à partir d'alignements multiples de Mlh1 eucaryotes et de MutL bactériens (**Figure 3 - 2(A)**). Cette modélisation suggère que Mlh1 présente deux grandes boucles additionnelles par rapport à ecMutL: la première est située entre l'hélice α_B et le feuillet β du domaine régulateur avec une hélice orientée différemment dans les deux modèles. La seconde boucle est située entre les hélices α_D et α_E et est prédite comme non structurée (**Figure 3 - 2(B)**).

Nous avons alors projeté sur la surface de ce modèle, le degré de conservation des résidus de Mlh1 à l'aide du programme ConSurf et des alignements multiples des séquences Mlh1 eucaryotes. Cette représentation montre que la surface du modèle du domaine C-terminal de Mlh1 présente deux zones particulièrement conservées (**Figure 3 - 2(C)**). La première se compose essentiellement de résidus du feuillet β et de l'hélice α_F à l'extrémité C-terminale. Ces éléments de structure secondaire correspondent au domaine de dimérisation proposé pour MutL d'*E. coli*. Ce domaine conservé est situé à proximité des 5 acides aminés, dont la position R547, identifiés par Argueso et al comme importants pour l'interaction avec Pms1. Cette première zone semble donc correspondre au site S1 d'interaction avec Pms1. La seconde zone conservée est située à plus de 40Å de ce premier site en plein cœur du domaine régulateur d'ecMutL. Cette zone est formée de résidus non contigus dans la séquence et est située autour des positions 578, 630 et 682. Cette zone comprend donc les résidus Y630 et E682 impliqués dans l'interaction avec Ntg2 et semble donc correspondre au site 2 d'interaction de Mlh1 avec les protéines contenant un motif MIP. Le troisième résidu identifié lors du 1er tour de mutagenèse, S513, n'a pas pu être modélisé car la structure du CTD d'ecMutL débute au résidu Q433, position équivalente au résidu F533 de la protéine Mlh1. Ce modèle permet de proposer une liste des acides aminés susceptibles de constituer le site S2 et notamment d'acides aminés proches de la position F578, zone non encore explorée.

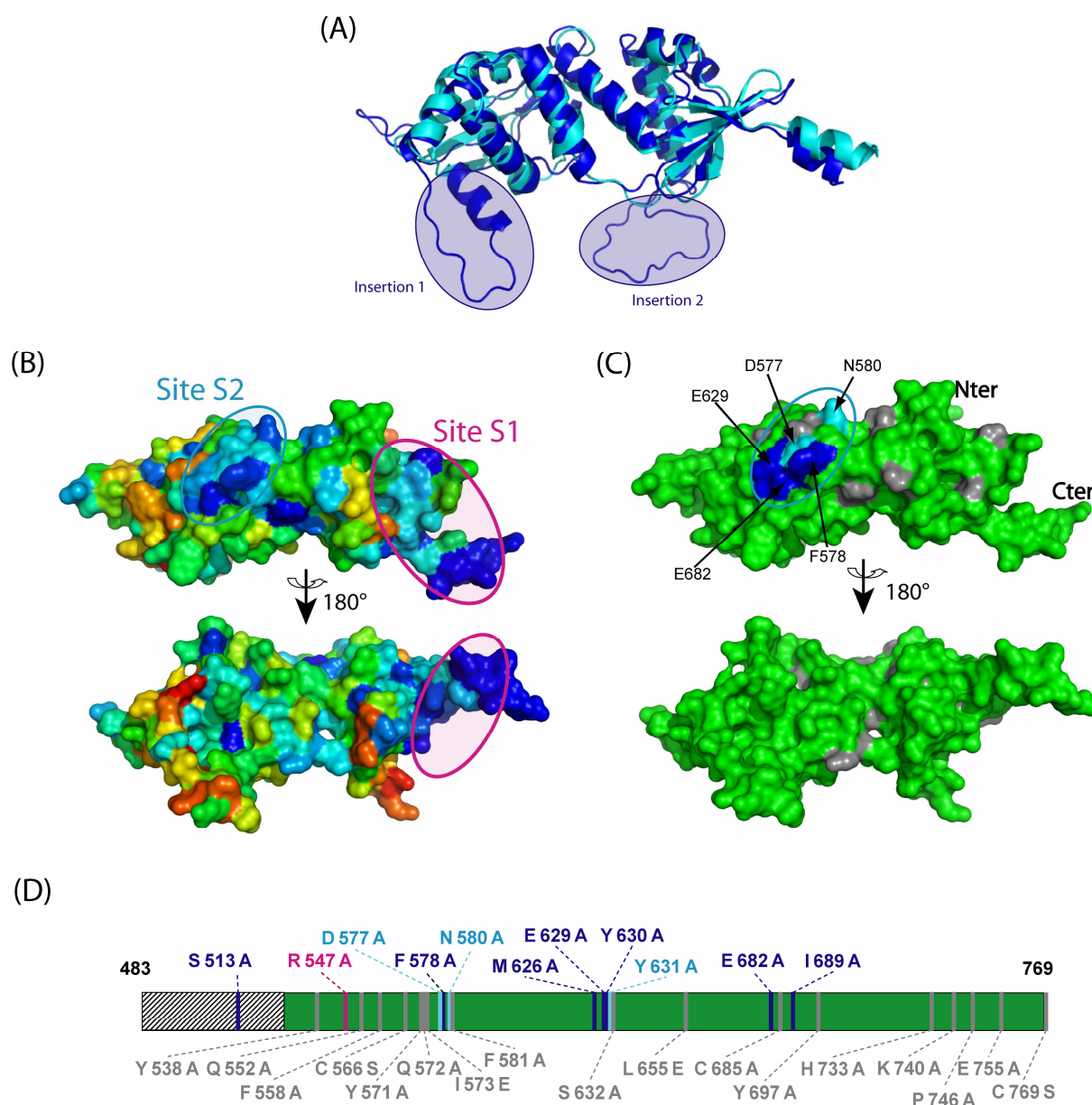


Figure 3 - 2 : Modélisation du domaine C-terminal de Mlh1 et localisation potentielle des sites S1 d'hétérodimérisation et S2 d'interaction avec les protéines contenant un motif MIP. (A) Superposition de la structure cristallographique du domaine C-terminal d'ecMutL (en cyan) et du modèle par homologie du domaine C-terminal de Mlh1 (en bleu) (B) Représentation de la surface du modèle du CTD de Mlh1 en fonction de la conservation des résidus (programme ConSurf) ; les résidus peu conservés sont en rouge, ceux très conservés sont en bleu (C) Représentation de la surface du modèle du CTD de Mlh1 en fonction des résultats de mutagenèse dirigée et double hybride ; les positions des mutations sans effet sur les interactions Pms1 et Ntg2 sont représentées en gris ; les positions des mutations qui annulent ou affaiblissent l'interaction avec Ntg2 sont marquées respectivement (-) et (+/-) et sont représentées en bleu et cyan respectivement. La position R547 (non visible car sur l'autre face de Mlh1) est située au niveau du site1 (D) Les 28 positions testées en double-hybride sont représentées sur la séquence de Mlh1 suivant le même code de couleur que ci-dessus ; la zone hachurée correspond à une zone non modélisée car absente de la structure cristallographique d'ecMutL

2nd série de mutants ponctuels sur la base de la modélisation par homologie de Mlh1 de levure : Une seconde série de 9 mutants de Mlh1 a alors été construite. Cette série est ciblée sur la zone conservée mise en évidence par la modélisation moléculaire (**Figure 3 - 2(D)**). Les expériences de double-hybride sur les mutants de la seconde série montrent que quatre d'entre eux (F578A, M626A, E629A et I689A) ne sont plus capables de lier Ntg2 alors qu'ils interagissent encore avec Pms1; trois autres mutants (D577, N580 et Y631) sont partiellement affectés dans leur interaction avec Ntg2 (**Tableau 3 - 2**). Les deux derniers mutants (S632A et Y697A) n'affectent ni l'interaction avec Pms1 ni celle avec Ntg2. Ce second tour de mutagenèse a donc permis d'identifier 7 nouvelles positions du site 2. Ce second tour a notamment permis de mettre en évidence le rôle d'acides aminés situés dans une nouvelle partie de la séquence de Mlh1 (D577, F578 et N580) dont l'implication avait été suggérée par la modélisation.

Tableau 3 - 2: Résultats du 2nd tour de mutagenèse sur le domaine C-terminal de Mlh1 et expériences de double-hybride pour l'interaction avec Pms1, Exo1 ou Ntg2; les interactions annotées (+) traduisent des colonies bleues, celles (+/-) traduisent des colonies bleues clair; celles (-) traduisent des colonies blanches; celles (n. d.) n'ont pas été déterminées. Les trois positions identifiées au 1^{er} tour sont rappelées en bas du tableau

Mutation ponctuelle du CTD de Mlh1	Interaction en double hybride		
	Pms1	Exo1	Ntg2
D577A	+	n. d.	+/-
F578A	+	-	-
N580A	+	n. d.	-
M626A	+	n. d.	-
E629A	+	n. d.	-
Y631A	+	n. d.	+/-
S632A	+	n. d.	+
I689A	+	n. d.	-
Y697A	+	n. d.	+
S513A	+/-	-	-
Y630A	+	-	-
E682A	+	-	-

Au total, ces deux tours de mutagenèse ont permis de proposer un ensemble de dix positions de Mlh1 impliquées dans l'interaction avec Ntg2 et susceptibles de constituer le site 2 de Mlh1. 4 parties différentes de la séquence de la région C-terminale de Mlh1 ont été identifiées: (i) près de S513 dans l'extrémité N-terminale non modélisée, (ii) autour des positions D577, F578 et N580, (iii) autour des positions M626, E629, Y630 et Y631 et enfin (iv) autour des positions E682 et I689 (**Figure 3 - 2(E)**). Nous avons alors choisi de confirmer ces résultats obtenus par des mesures

d'interactions directes par calorimétrie de la région C-terminale de Mlh1 et des peptides dérivés des protéines Ntg2, Exo1 et Sgs1 et contenant le motif MIP de ces protéines.

3.1.2 PURIFICATION ET CARACTERISATION BIOPHYSIQUE DE LA REGION C-TERMINALE DE MLH1

Les domaines C-terminaux des protéines sauvages Mlh1 de levure (notée Mlh1(CTD) par la suite) et MLH1 humaine (notée MLH1(CTD) par la suite) ainsi que les variants E682A et R547A de la protéine de levure sont produits et purifiés suivant un protocole identique³.

Le domaine C-terminal de Mlh1 est surexprimé en fusion avec en position N-terminale la Glutathion-S-Transférase (GST) étiquetée 6×His. La protéine de fusion possède un site de clivage à la protéase TEV entre l'étiquette 6×His-GST et Mlh1(CTD) (485 à 769). Les extraits bruts, avant et après induction des cultures du domaine C-terminal de Mlh1, sont analysés sur gel SDS-PAGE. Le gel indique une forte surexpression d'une bande à la hauteur de la protéine de fusion GST-Mlh1 (PM 62 kDa) (résultat non montré). Les surnageants après lyse cellulaire et centrifugation de 1L de culture d'*E. coli*, sont chargés sur une colonne de glutathion (**Figure 3 - 3(A)**). Environ 100mg de GST- Mlh1(CTD) natif et 20mg des fusions des mutants de Mlh1 par litre de culture, sont élués de la colonne d'affinité (les concentrations sont estimées par la méthode de Bradford). Environ 400mg de fusion GST-domaine C-terminal humain MLH1 (486 à 756) sont obtenus par litre de culture.

L'éluat de la colonne d'affinité est ensuite coupé avec 1% (m/m) (pour masse/masse) de protéase TEV à 4°C sur la nuit. Les protéines portant une étiquette 6×His sont éliminées par passage sur une colonne de nickel. Mlh1(CTD) ou MLH1(CTD) ne sont pas retenus (**Figure 3 - 3(A)**). Les protéines humaines et de levure ont des rendements différents à cette étape. La moitié environ des protéines de fusion correspond à la GST qui est fixée sur la colonne. Sur l'autre moitié, on récupère en sortie directe environ 60% de Mlh1(CTD) de levure attendu et 30% de MLH1(CTD) humain attendu. Pour MLH1(CTD) humain, ce rendement de purification a été amélioré en coupant directement sur la colonne de glutathion la protéine de fusion GST-MLH1.

La protéine Mlh1(CTD) est ensuite concentrée et purifiée sur colonne par exclusion de taille. Le chromatogramme de Mlh1(CTD) de levure ou humain, présente un premier pic (de 1 à 10% de l'aire totale à 280nm) autour du volume mort de la colonne. Le second pic, pic principal, (environ 80% de l'aire totale à 280nm) sort à un volume d'élution de 132ml. Ce volume d'élution est légèrement supérieur au

³ Le détail des protocoles de purification est donné en ANNEXE 3 & ANNEXE 4

volume de sortie de l'Albumine (PM=67kDa). Le chromatogramme du Mlh1(CTD) de levure présente, en général, un 3^{ème} pic (environ 10% de l'aire totale à 280nm) vers 152ml. Ce volume d'élution est légèrement supérieur à celui observé avec l'Ovalbumine (PM=43kDa) (**Figure 3 - 3(B)**). Ce pic n'est pas présent sur le chromatogramme de MLH1(CTD) humain (**Figure 3 - 3(C)**). Cette étape de polissage élimine de légers contaminants d'après le gel SDS-PAGE et présente un rendement de 90% pour les différents CTD de Mlh1 purifiés.

L'analyse sur gel SDS-PAGE montre que les deux ou trois pics, selon les cas, contiennent essentiellement Mlh1(CTD). Le domaine C-terminal de Mlh1 a un poids moléculaire de 31,2 kDa chez la levure et 31,4 kDa chez l'humain. Le premier pic correspond à une forme agrégée de la protéine. Le second pic, qui contient l'essentiel Mlh1(CTD), possède un rayon de Stokes qui, d'après les calibrations, correspond à un PM apparent de 64 kDa. Le troisième pic correspond également à Mlh1(CTD) sur gel avec un poids moléculaire apparent de 35kDa. Ces résultats suggèrent que le second et le troisième pic de cette étape de purification correspondent respectivement à une forme dimérique et monomérique de Mlh1(CTD). Les chromatogrammes obtenus pour les variants R547A et E682A lors du passage sur cette même colonne d'exclusion de taille sont similaires, permettant, au final, l'obtention d'un échantillon de qualité satisfaisante (**Figure 3 - 3(D)**). A l'issue de cette étape de polissage, sont obtenus près de 20mg de Mlh1(CTD) natif de levure et 5mg des deux variants de levure, R547A et E682A, par litre de culture d'*E. coli*. On obtient environ 50mg de MLH1(CTD) natif humain.

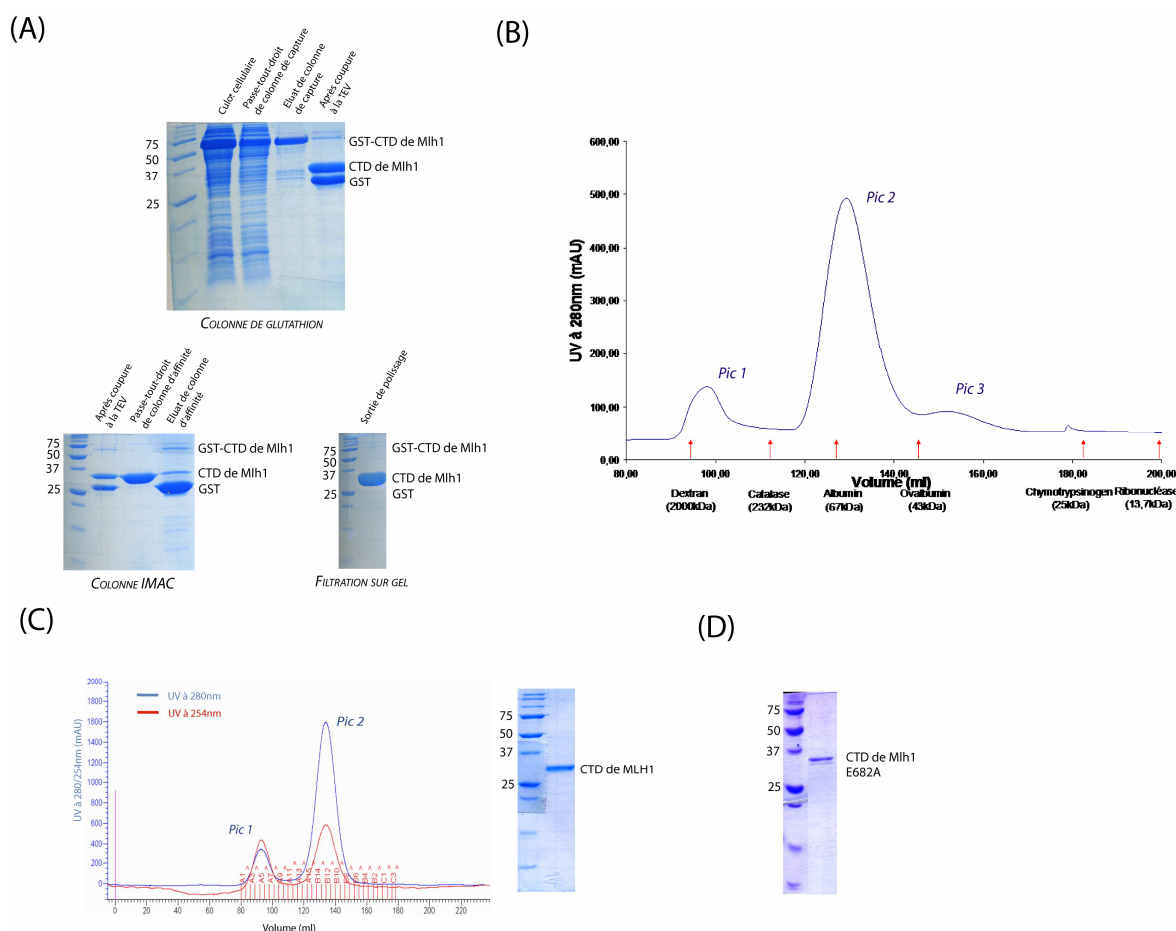


Figure 3 - 3: Gels SDS-PAGE et chromatogrammes de différentes étapes de purification du domaine C-terminal de Mlh1 de levure et humain (A) Gels SDS-PAGE des trois étapes de purification de Mlh1(CTD) natif (B) Chromatogramme de Mlh1(CTD) de levure natif sur colonne d'exclusion de taille (C) Chromatogramme de MLH1(CTD) humain natif sur colonne d'exclusion de taille et gel SDS-PAGE du pic principal. (D) Gel SDS-PAGE en sortie de chromatographie d'exclusion stérique du pic principal du domaine C-terminal de Mlh1 E682A

Pour s'assurer de la qualité des échantillons obtenus en fin de purification, les protéines purifiées sont analysées par des expériences de dichroïsme circulaire et de diffusion dynamique de la lumière (DLS).

Les spectres de dichroïsme circulaire de Mlh1(CTD) de levure (**Figure 3 - 4(A)**) et humain (**Figure 3 - 4(D)**) présentent, tous les deux, les caractéristiques d'un spectre de protéines repliées. Les spectres obtenus présentent les caractéristiques d'un repliement α/β . On observe en effet un signal caractéristique d'un mélange d'hélices α (maximum vers 195nm et minima à 208 et 222nm) et de feuillets β (maximum vers 195nm et minimum à 215nm). La dénaturation thermique suivie par dichroïsme circulaire du domaine C-terminal de Mlh1 de levure montre un signal de déstructuration vers 54°C (**Figure 3 - 4(B)**). La transition intervient à 44°C pour le domaine humain (**Figure 3 - 4(E)**). Le variant E682A de levure présente des caractéristiques identiques à la protéine native⁴.

⁴ Le spectre CD et la dénaturation thermique du variant E682A sont visibles en ANNEXE 5

L'analyse par diffusion dynamique de la lumière montre que les échantillons en sortie de purification sont monodisperses, et ce, jusqu'à deux semaines après la production. Au delà, on observe la formation d'agrégats (mis en évidence en DLS par l'apparition d'une population de taille supérieure à 100nm). Au-delà de deux semaines de conservation à 4°C, on observe également une légère dégradation de la protéine (bandes situées sous la bande principale sur gel SDS-Page). Les CTD de Mlh1 et MLH1 ont un rayon de Stokes apparent de 6.97 et 6.56nm respectivement, ce qui correspond à des protéines globulaires d'environ 64 et 54kDa (**Figure 3 - 4(C)&(F)**). Ce résultat est en accord avec les volumes d'élution mesurés sur colonne par exclusion de taille. Ceci suggère qu'à la concentration de 0.5mg/ml en Mlh1(CTD) ou MLH1(CTD), les deux protéines se trouvent vraisemblablement dans un état homodimérique.

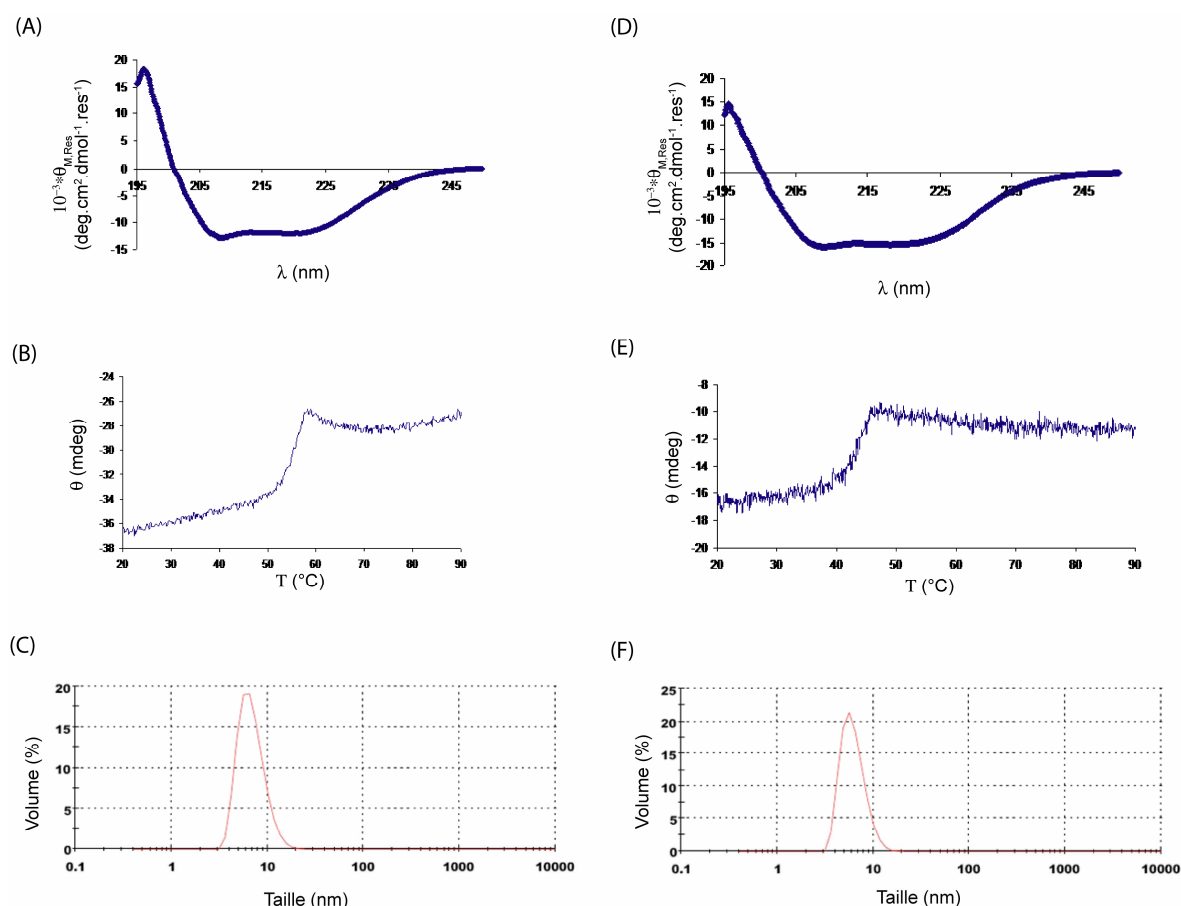


Figure 3 - 4 : Analyses par dichroïsme circulaire et diffusion de lumière du domaine C-terminal de Mlh1
 (A) Spectre de dichroïsme circulaire du domaine C-terminal de Mlh1 ; le spectre est enregistré entre 195 et 250nm à 20°C ; l'échantillon est à une concentration de 7 μ M (B) Dénaturation thermique du domaine C-terminal de Mlh1 ; la dénaturation est suivie par dichroïsme circulaire à 222nm entre 20 et 90°C (C) Spectre de diffusion dynamique de lumière du domaine C-terminal de Mlh1 pondéré en volume ; l'échantillon est à une concentration de 0.5mg/ml (D) Spectre de dichroïsme circulaire du domaine C-terminal de MLH1 ; le spectre est enregistré entre 195 et 250nm à 20°C ; l'échantillon est à une concentration de 10 μ M (E) Dénaturation thermique du domaine C-terminal de MLH1 ; la dénaturation est suivie par dichroïsme circulaire à 222nm entre 20 et 90°C (F) Spectre de diffusion dynamique de lumière du domaine C-terminal de MLH1 pondéré en volume ; l'échantillon est à une concentration de 0.5mg/ml

A l'issue de la purification, ces résultats indiquent que les protéines ainsi purifiées, sont correctement repliées et présentent une bonne homogénéité de taille et forme.

3.1.3 CARACTERISATION PAR CALORIMETRIE DES INTERACTIONS DE PEPTIDES CONTENANT UN MOTIF MIP AVEC LA REGION C-TERMINALE DE MLH1

i. Spécificité de l'interaction du site S2 du domaine C-terminal de Mlh1 avec Ntg2, Exo1 et Sgs1

Afin d'évaluer l'affinité entre le site S2 de Mlh1 et ses partenaires, nous avons choisi de caractériser par microcalorimétrie de titration isotherme, appelée calorimétrie par la suite, les interactions entre le domaine C-terminal de Mlh1 purifié et des peptides contenant un motif MIP. Le choix de l'utilisation de peptides à la place des protéines partenaires entières a été motivé par deux éléments: (i) les motifs MIP sont situés dans des régions prédites comme non structurées des protéines partenaires de Mlh1. (ii) aucun protocole satisfaisant de purification des protéines Ntg2, Exo1 et Sgs1 entières n'est décrit. Les essais de purification au laboratoire de la protéine Ntg2 montrent que la forme principale purifiée est une forme tronquée au niveau de son extrémité N-terminale et qui ne contient plus le motif MIP. Les articles décrivant la purification de la protéine Exo1 montrent que l'extrémité C-terminale de Exo1 qui contient le motif MIP est clivée lors de la purification (Fiorentini et al., 1997). Dans un premier temps nous avons donc choisi de mesurer les interactions entre la région C-terminale de Mlh1 de levure et des peptides de 12 à 13 acides aminés centrés sur le motif MIP des protéines Ntg2, Exo1 et Sgs1.

Les thermogrammes obtenus en titrant Mlh1(CTD) par les peptides pNtg2(13-mère), pSgs1(12-mère) et pExo1(13-mère) présentent un signal d'échange de chaleur exothermique (l'enthalpie de la réaction est négative) (**Figure 3 - 5(A-C) & Tableau 3 - 3**). L'échange de chaleur observé n'est pas mesuré lorsque l'on injecte le peptide seul dans le tampon de dialyse de Mlh1(CTD) ou inversement, lorsque l'on titre la protéine par son tampon de dialyse mis dans la seringue (données non montrées). Ceci souligne que l'échange de chaleur provient de l'interaction entre le CTD de Mlh1 et les peptides contenant le motif MIP. L'intégration de l'aire comprise sous les pics du thermogramme donne accès à l'isotherme de titration de l'interaction. Pour les trois peptides, l'isotherme obtenue peut être modélisée par un modèle avec un seul site de fixation. Les courbes d'ajustement obtenues pour les isothermes indiquent que les peptides pNtg2, pSgs1 et pExo1 ont une affinité similaire avec des constantes de dissociation égales à $9.0 \pm 0.6 \mu\text{M}$, $9.3 \pm 2.9 \mu\text{M}$ et $8.1 \pm 1.4 \mu\text{M}$, respectivement.

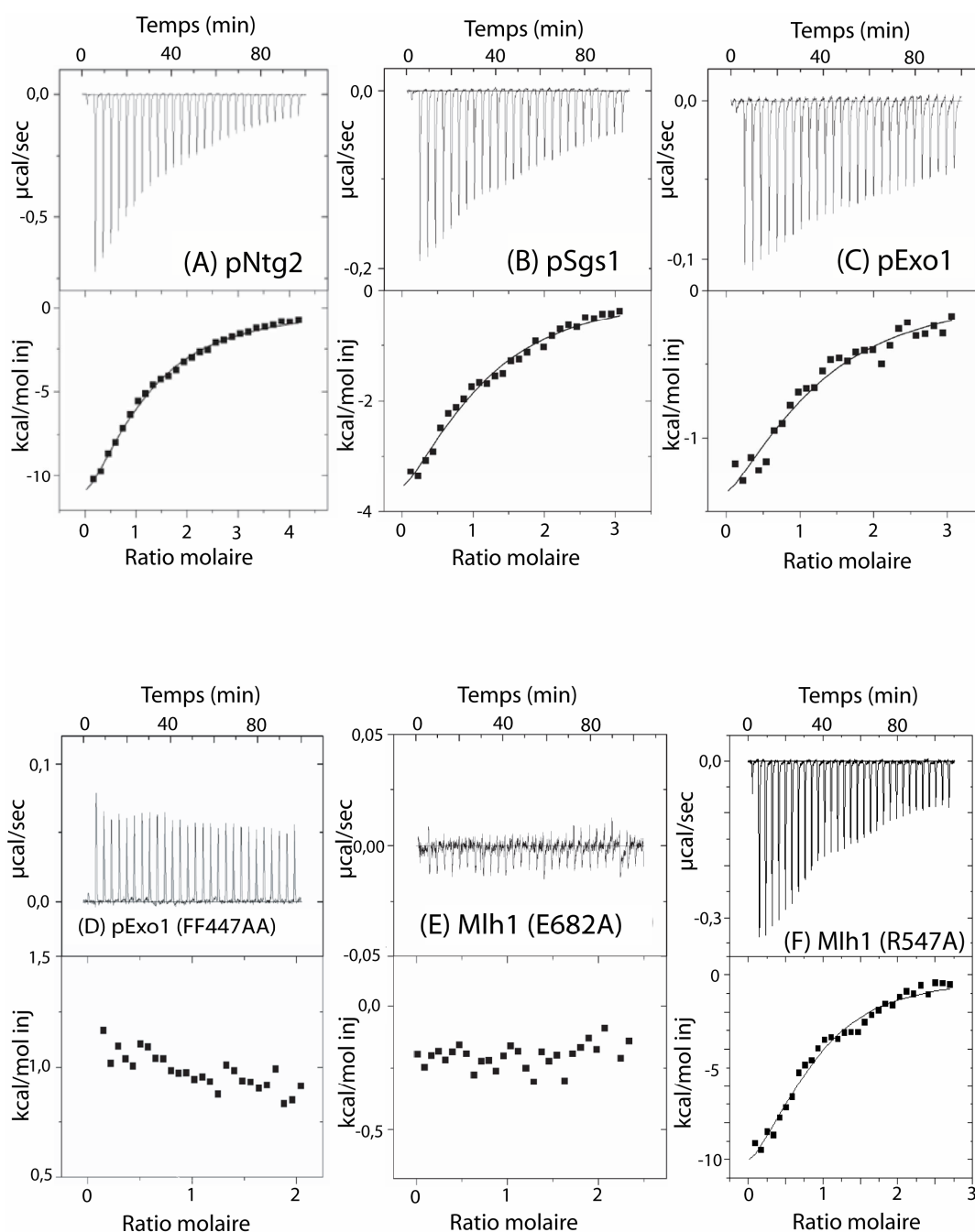


Figure 3 - 5 : Mesures par microcalorimétrie de l'interaction entre le domaine C-terminal de Mlh1 et des peptides contenant le motif MIP ; les expériences sont réalisées à 30°C, la protéine est à 15 μM , le peptide est à 150 μM ; les paramètres thermodynamiques sont présentés dans le Tableau 3 - 3 (A) CTD de Mlh1 natif et pNtg2(13mère) (B) CTD de Mlh1 natif et pSgs1(12mère) (C) CTD de Mlh1 natif et pExo1(13mère) (D) CTD de Mlh1 natif et pExo1(FF447AA) (E) CTD de Mlh1 E682A et pNtg2(13mère) (F) CTD de Mlh1 R547A et pNtg2(13mère)

Pour confirmer la spécificité de l'interaction, nous avons testé un peptide du motif MIP d'Exo1 contenant une mutation en alanine sur les deux résidus aromatiques, F447 et F448, identifiés comme importants pour l'interaction par double hybride (Gellon et al., 2002). Le thermogramme résultant de l'injection du peptide pExo1-F447A-F448A présente des pics de faible intensité, constants au cours des injections. Ceci met en évidence qu'il n'y a pas d'interaction entre ce peptide et

Mlh1(CTD) (**Figure 3 - 5(D) & Tableau 3 - 3**) et souligne la spécificité de l'interaction observée avec le peptide pExo1 sauvage. La même observation est obtenue avec le peptide pNtg2(13-mère) S24A, qui présente une mutation en alanine sur la sérine du motif MIP (**Tableau 3 - 3**).

Nous avons alors analysé du côté de Mlh1, l'interaction du peptide pNtg2(13-mère) avec les mutants E682A et R547A de Mlh1. Nous n'observons aucune interaction par calorimétrie entre le peptide pNtg2(13-mère) et le mutant du site S2, E682A. Ce résultat confirme les données de double-hybride sur le rôle de cet acide aminé dans l'interaction avec Ntg2 (**Figure 3 - 5(E) & Tableau 3 - 3**). Le variant R547A du site S1 de Mlh1, impliqué dans l'hétérodimérisation avec Pms1, donne un thermogramme avec ce même peptide pNtg2 proche de celui observé la protéine native (**Figure 3 - 5(F) & Tableau 3 - 3**). Ce résultat est également en bon accord avec les résultats obtenus par double hybride sur le variant R547A de Mlh1.

Ces premières mesures de calorimétrie indiquent clairement une interaction directe et spécifique entre les motifs MIP des protéines Ntg2, Exo1 et Sgs1 et le site S2 de Mlh1. Les valeurs des constantes de dissociation obtenues, de l'ordre du micromolaire, situent cette interaction parmi les interactions protéines-protéines ou protéines-peptides modestes (Aloy and Russell, 2006).

On peut noter que les trois peptides pNtg2, pExo1 et pSgs1 interagissent avec la protéine Mlh1 selon une même affinité et donc une même variation d'enthalpie libre ΔG° . Néanmoins, ces peptides présentent des variations d'entropie et d'enthalpie différentes. L'interaction avec le peptide pNtg2 est conduite par l'enthalpie, (variation enthalpique négative donc favorable). Cette enthalpie favorable est partiellement compensée par une variation entropique défavorable. Ce mode d'interaction est souvent associé à de nombreux contacts de Van der Waals et à la formation de liaisons hydrogène. A l'opposé, l'interaction avec le peptide pExo1 est conduite par l'entropie (variation entropique négative donc favorable). La variation enthalpique est près de 10 fois plus faible que celle observée avec le peptide pNtg2. Ce mode d'interaction est souvent associé à des interactions avec des contributions importantes d'acides aminés hydrophobes et/ou des changements de conformation (Holdgate, 2001). Les variations thermodynamiques observées avec le peptide Sgs1 représentent un cas de figure intermédiaire entre le peptide Ntg2 et le peptide Exo1. L'interaction est conduite par une enthalpie plus faible que le peptide Ntg2 et la variation entropique est presque nulle. Il est difficile à cette étape d'en déduire des enseignements précis sur le mode d'action de ces trois peptides; néanmoins ces différences suggèrent l'existence de modes d'interaction partiellement différents. La présence du motif MIP dans ces trois peptides suggère que ces différences viennent des interactions réalisées par les acides aminés situés de part et d'autre du motif MIP.

Tableau 3 - 3 : Analyses thermodynamiques par calorimétrie à titrage isotherme de l'interaction entre la région C-terminale de Mlh1 et des peptides de Ntg2, Exo1 et Sgs1

Protéine	Peptide	Séquence	ΔH° (kcal/mol)	K_a ($10^5 M^{-1}$)	K_d (μM)	ΔG° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol)
Mlh1 native	pNtg2	²⁰ VEVRSKYFKKNER	-20.5±1.5	1.11±0.08	9.0±0.6	-7.00	+13.5
Mlh1 native	pExo1	⁴⁴¹ KDTRSKFFNKPS	-2.6±0.7	1.07±0.34	9.3±2.9	-7.00	-4.40
Mlh1 native	pSgs1	¹³⁷⁷ QSRGKSKFFGAN	-6.5±0.9	1.24±0.21	8.1±1.4	-7.05	-0.55
Mlh1 R547A	pNtg2		-18.8±1.5	1.90±0.20	5.3±0.5	-7.30	+11.5
Mlh1 E682A	pNtg2			Pas d'interaction			
Mlh1 native	pExo1 (FF447AA)	⁴⁴¹ KDTRSK AA NKPS		Pas d'interaction			
Mlh1 native ⁵	pNtg2 (S24A)	²⁰ VEVR AK YFKKNER		Pas d'interaction			

En guise de comparaison et afin de confirmer la mesure d'affinité par une autre méthode biophysique, l'affinité de l'interaction entre Mlh1(CTD) et le peptide pNtg2(13-mère) a été mesurée par interférométrie par rétrodiffusion (BSI pour Back Scattering Interferometry) (Bornhop et al., 2007). Dans le cadre d'une collaboration avec la start-up Molecular Sensing, (Montara, USA) qui développe cette nouvelle technologie, nous avons pu obtenir des courbes de titration par BSI et mesurer une constante de dissociation de $4.60 \mu M \pm 0.53^6$. Cette valeur est en bon accord avec celle obtenue par calorimétrie. Le peptide d'Exo1-F448A-F449A, contenant la double mutation sur le motif MIP, ne présente pas de signal d'interaction par BSI.

ii. Conservation de l'interaction du site S2 du domaine C-terminal de MLH1

Il a été évoqué précédemment que le motif MIP d'Exo1 était conservé, notamment chez l'humain, et que ce même motif se retrouvait également chez l'hélicase BLM, un des homologues de Sgs1 chez l'homme.

Une étude en double-hybride a montré que le domaine C-terminal de BLM (1034-1417) qui contient le motif MIP, interagit avec la région C-terminale de MLH1 (Pedrazzi et al., 2001). Nous avons donc utilisé la même stratégie que celle mise en œuvre avec les protéines de levure. Ainsi, nous avons analysé l'interaction entre MLH1(CTD) humain et des peptides dérivés des motifs MIP des protéines humaines, BLM et d'EXO1: dans les deux cas, nous observons une interaction avec une affinité de l'ordre de la dizaine de micromolaire, à savoir $9.0 \mu M$ avec le peptide pBLM(10-mère) et $17.4 \mu M$ avec le peptide pEXO1(16-mère) (**Figure 3 - 6 & Tableau 3 - 4**). En revanche, aucune interaction n'est observée lorsque les sérines S1323 de BLM ou S504 d'EXO1 sont mutées en alanine (**Tableau 3 - 4**). La même conclusion est obtenue avec le peptide MIP d'EXO1 contenant la mutation FF506AA (**Tableau 3 - 4**). Ces

⁵ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6– (A)

⁶ Les courbes expérimentales de BSI sont présentées en ANNEXE 7

thermogrammes montrent la conservation chez l'homme des interactions entre la région C-terminale de MLH1 et des peptides contenant le motif MIP dérivés des protéines BLM et EXO1.

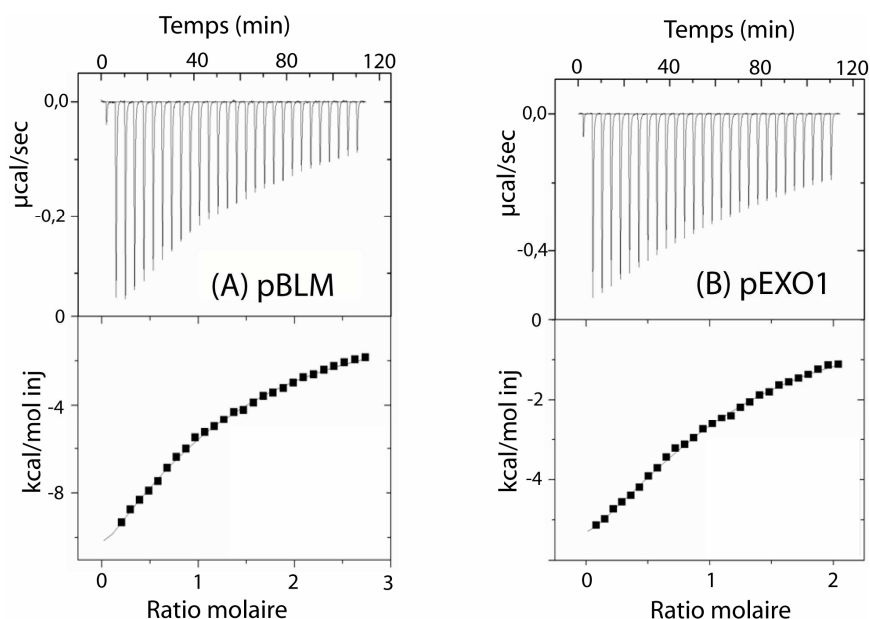


Figure 3 - 6 : Mesures par calorimétrie des interactions entre le domaine C-terminal de MLH1 et des peptides contenant le motif MIP ; les expériences sont réalisées à 30°C, la protéine étant à 15µM, le peptide à 150µM ; les paramètres thermodynamiques sont présentés dans le Tableau 3 - 4 (A) CTD de MLH1 natif et pBLM(10-mère) (B) CTD de MLH1 natif et pEXO1(16-mère)

Tableau 3 - 4 : Mesures par calorimétrie de l'interaction entre la région C-terminale de MLH1 et des peptides de BLM, EXO1 et Ntg2 ; entre la région C-terminale de Mlh1 et le peptide de BLM

Protéine	Peptide	Séquence	ΔH° (kcal/mol)	K_a ($10^5 M^{-1}$)	K_d (µM)	ΔG° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol)
MLH1 native	pBLM	¹³²¹ VSSH Y FASKT	-21.0±1.3	1.11±0.07	9.0±1.2	-7.00	+14.0
MLH1 native	pEXO1	⁴⁹⁹ VVG T RSR F CCSSDST	-8.5±0.2	0.57±0.03	17.4±0.9	-6.60	+1.90
MLH1 native	pNtg2	²⁰ VEVRSKY F KKNER	-22.1±0.3	10.9±1.5	0.92±0.01	-8.40	+13.7
Mlh1 native ⁷	pBLM	¹³²¹ VSSH Y FASKT	-2.2±0.2	1.20±0.14	8.3±1.0	-7.05	-4.85
MLH1 native ⁸	pBLM (S1323A)	¹³²¹ VSAHYFASKT			Pas d'interaction		
MLH1 native ⁹	pEXO1 (S504A)	⁴⁹⁹ VVG T RSR F CCSSDST			Pas d'interaction		
MLH1 native ¹⁰	pEXO1 (FF506AA)	⁴⁹⁹ VVG T RSRA A CCSSDST			Pas d'interaction		

A cette étape, nous avons choisi d'analyser par dichroïsme circulaire (ou CD pour Circular Dichroism) le complexe MLH1 humain/pNtg2 avec une stoechiométrie 1:1 et à une concentration de 20µM. Dans ces conditions, étant donnée la constante

⁷ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (E)

⁸ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (B)

⁹ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (C)

¹⁰ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (D)

de dissociation mesurée pour ce complexe (K_d de $0.92\mu\text{M}$), le CTD de MLH1 et le peptide pNtg2(13-mère) sont, très majoritairement, sous forme liée. Les spectres de dénaturation thermique montrent que la présence du peptide ne s'accompagne que d'une très légère stabilisation du CTD de MLH1. La température de dénaturation thermique passe de 44°C pour la protéine seule à 46°C pour le complexe protéine peptide. Le même spectre est effectué avec le CTD de MLH1 et le peptide pEXO1(16-mère) qui présente une constante de dissociation de $17.4\mu\text{M}$ avec MLH1. Dans les conditions utilisées en dichroïsme circulaire, une fraction significative de la protéine est vraisemblablement non liée. Nous observons dans ce cas une température de dénaturation identique à celle du CTD de MLH1 seul, *i.e.* 44°C .

Afin de vérifier si l'interaction avec le peptide est accompagnée par des changements de structure secondaire majeurs du peptide ou de la protéine MLH1, nous avons effectué les spectres de dichroïsme circulaire à une concentration de $40\mu\text{M}$ en utilisant une cuve dont le chemin optique est dix fois plus petit ($0,1\text{mm}$ contre 1mm dans nos essais standard). Cette cuve permet d'augmenter la concentration en protéine sans saturer le photomultiplicateur de l'appareil. Les spectres effectués entre 195 et 260nm , montrent que le repliement global est identique entre la protéine seule et la protéine liée au peptide pNtg2(13-mère). Le spectre effectué avec le peptide pEXO1(16-mère) ne montre pas de modification majeure de structure secondaire alors qu'à cette concentration, la forme liée devient majoritaire. Les peptides pour leur part, ont un spectre caractéristique de peptides non structurés (données non montrées). Ceci permet de conclure que le CTD de MLH1 ou que le peptide ne subit pas un important réarrangement structural comme cela a pu être observé par CD pour d'autres interactions protéine-peptide (Popescu et al., 2003).

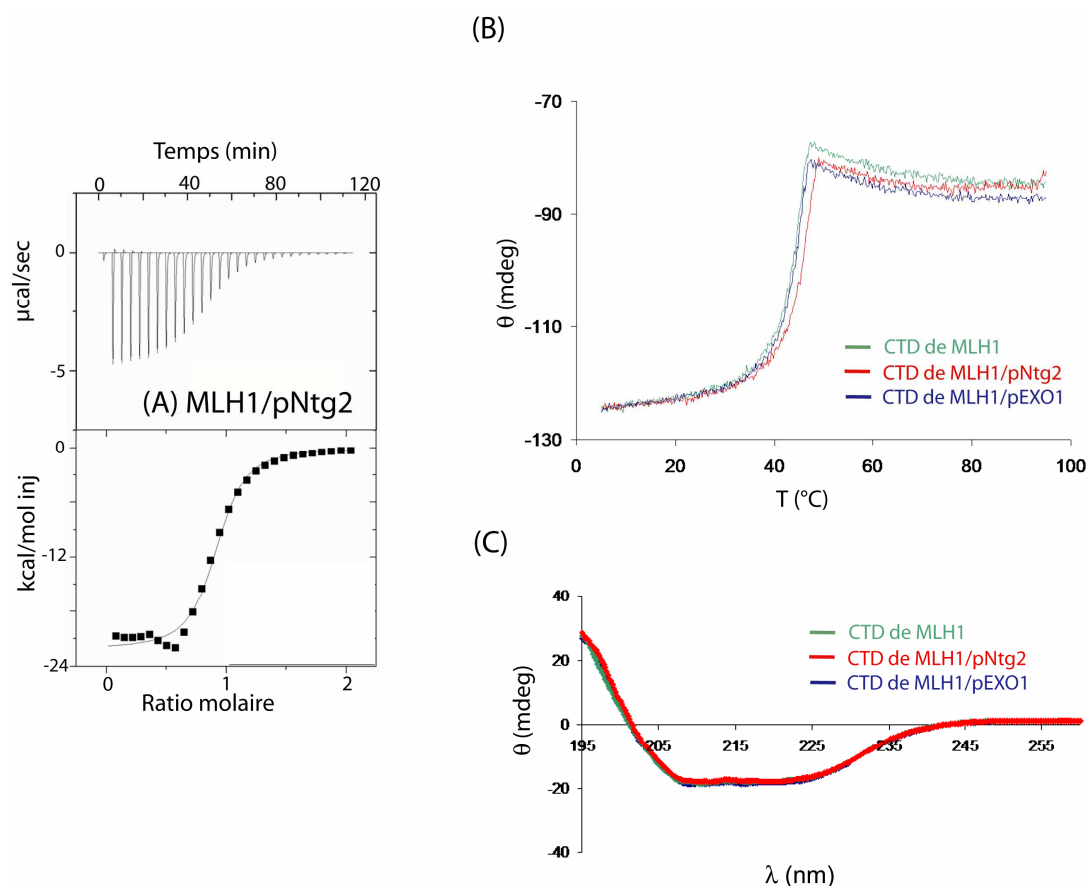


Figure 3 - 7 : Interactions entre la région C-terminale de MLH1 humain et les peptides Ntg2 de levure ou EXO1 humain (A) Thermogramme et isotherme de titration du CTD de MLH1 natif et pNtg2(13-mère) ; les expériences sont réalisées à 30°C, la protéine étant à 60μM, le peptide à 600μM ; les paramètres thermodynamiques sont présentés dans le Tableau 3 - 4 (B) Courbes de dénaturation thermique suivie par dichroïsme circulaire du CTD de MLH1 seul et en présence équimolaire des peptides pNtg2(13-mère) ou pEXO1(16-mère) ; la dénaturation est suivie par dichroïsme circulaire à 222nm entre 5 et 95°C ; la protéine est à une concentration de 20μM (C) Spectre de dichroïsme circulaire du domaine C-terminal de MLH1 seul et en présence équimolaire de pNtg2(13-mère) et pEXO1(16-mère) ; le spectre est enregistré entre 195 et 260nm à 20°C ; l'échantillon est à une concentration de 40μM

iii. Mutation en alanine du motif MIP du peptide Ntg2

Lors de l'étude initiale du motif MIP de Ntg2 (Gellon et al., 2002), des tests de double hybride avaient été effectués pour évaluer l'impact sur l'interaction avec Mlh1 d'une mutation en alanine des 5 acides aminés, RSKYF, du motif MIP de Ntg2. Afin de valider par une autre méthode ces résultats et de préciser le rôle des différentes positions centrales du peptide Ntg2 pour l'interaction avec la région C-terminale de Mlh1, nous avons décidé de mesurer l'interaction par calorimétrie de 5 peptides présentant une simple mutation en alanine (Tableau 3 - 5). Comme évoqué un peu plus haut, les mesures de calorimétrie confirment que la sérine est essentielle pour l'interaction. Les mesures réalisées avec les peptides pNtg2-Y26A et pNtg2-F27A montrent que ces deux positions annulent l'interaction avec le CTD de Mlh1. La mutation des résidus R23 et K25 diminue légèrement l'interaction avec Mlh1. Ces résultats sont en bon accord avec les résultats de double hybride et permettent par rapport au double-hybride de quantifier le rôle des acides aminés R23 et K25.

L'analyse des sites potentiels de phosphorylation de la séquence de Ntg2 sur le serveur de prédiction de ces sites chez la levure, NetPhosYeast, suggère que la serine S24 de Ntg2 est un site potentiel de phosphorylation. Nous avons analysé l'interaction d'un peptide synthétisé avec une serine phosphorylée sur la position 24. Ce peptide phosphorylé n'interagit plus avec la protéine Mlh1. Ce résultat confirme, en premier, la participation de la serine dans l'interaction. En ce qui concerne une éventuelle régulation négative de l'interaction avec la protéine Mlh1 par un mécanisme de phosphorylation, l'étape suivante est d'évaluer la présence d'une phosphorylation sur la serine du motif MIP des protéines Ntg2, Exo1 ou Sgs1.

Tableau 3 - 5 : Analyses thermodynamiques par calorimétrie entre la région C-terminale de Mlh1 et des peptides de Ntg2 ; analyse des mutants du motif MIP et d'un peptide phosphorylé

Protéine	Peptide	Séquence	ΔH° (kcal/mol)	K_a ($10^5 M^{-1}$)	K_d (μM)	ΔG° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol)
Mlh1 native	pNtg2	²⁰ VEV A SKYFKKNER	-40.8±0.4	3.09±0.07	3.24±0.8	-7.60	+33.2
Mlh1 native ¹¹	pNtg2 (R23A)	²⁰ VEV A SKYFKKNER	-36.2±1.3	0.90±0.03	11.2±0.5	-6.20	+30.0
Mlh1 native	pNtg2 (S24A)	²⁰ VEVR A KYFKKNER			Pas d'interaction		
Mlh1 native ¹²	pNtg2 (K25A)	²⁰ VEVR S A A YFKKNER	-31.4±0.8	0.70±0.01	14.2±0.4	-7.43	+24.0
Mlh1 native ¹³	pNtg2 (Y26A)	²⁰ VEVR S K A YFKKNER			Pas d'interaction		
Mlh1 native ^{14,15}	pNtg2 (F27A)	²⁰ VEVR S KY A KKNER			Pas d'interaction		
Mlh1 native ¹⁶	pNtg2 (S24S[P])	²⁰ VEVR S [P]KYFKKNER			Pas d'interaction		

iv. Influences de différents paramètres sur la force de l'interaction Mlh1(CTD)-peptide MIP: longueur du peptide, température et concentration en sels

Afin de préciser le mode d'interaction des peptides contenant un motif MIP avec la région C-terminale de Mlh1, nous avons décidé d'analyser par calorimétrie cette interaction en faisant varier trois paramètres :

- la longueur du peptide de part et d'autre du motif MIP
- la température de mesure de l'interaction
- la concentration en sels dans le tampon de mesure

¹¹ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (O)

¹² Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (P)

¹³ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (Q)

¹⁴ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (R)

¹⁵ Le thermogramme et l'isotherme du peptide dilué dans le tampon de dialyse sont visibles en ANNEXE 6 – (S)

¹⁶ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (T)

Influence de la longueur du peptide de part et d'autre du motif MIP : Dans le but d'identifier la longueur minimum des peptides Ntg2 et Exo1 pour l'interaction avec Mlh1, nous avons analysé l'affinité de peptides de longueurs variées dérivés des séquences de ces deux protéines (**Tableau 3 - 6**). Trois peptides Exo1 d'une longueur respective de 8, 12 et 16 acides aminés ont été analysés. Les peptides de 8 et 12 acides aminés présentent la même enthalpie et la même affinité. Le peptide Exo1 de 16 résidus est 1.6 fois plus affin. Trois peptides Ntg2 d'une longueur de 6, 13 et 16 acides aminés ont été analysés. Le peptide Ntg2 de 16 résidus a une affinité similaire à celui de 13 résidus. Le peptide pNtg2(6-mère) se lie à Mlh1(CTD) beaucoup plus faiblement. A la concentration où est réalisé le thermogramme (15µM), il n'est pas possible de mesurer précisément l'affinité mais on peut estimer une perte d'affinité d'un facteur au moins 20 ($K_d > 180\mu\text{M}$). Pour confirmer ce résultat, la mesure de l'interaction Mlh1(CTD)-pNtg2(6-mère) a été réalisée par la méthode BSI évoquée précédemment. En accord avec la calorimétrie, cette méthode suggère une perte d'affinité d'un facteur au moins 15 par rapport au le peptide de 13 acides aminés ($K_d > 65\mu\text{M}$ pour pNtg2(6-mère) contre 4µM pour pNtg2(13-mère)¹⁷).

Tableau 3 - 6 : Analyses thermodynamiques par calorimétrie à titrage isotherme de l'interaction entre la région C-terminale de Mlh1 et des peptides de Ntg2 et Exo1 en fonction de la longueur du peptide

Protéine	Peptide	Séquence	ΔH° (kcal/mol)	K_a (10^5M^{-1})	K_d (µM)	ΔG° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol)
Mlh1 native ¹⁸	pExo1 (8-mère)	⁴⁴³ TRSKFFNK	-2.8±0.2	1.14±0.34	8.8±1.5	-6.99	-4.16
Mlh1 native	pExo1 (12-mère)	⁴⁴¹ KDTRSKFFNKPS	-2.6±0.7	1.07±0.34	9.3±2.9	-7.00	-4.40
Mlh1 native ¹⁹	pExo1 (16-mère)	⁴³⁹ TLKDTRSKFFNKPSMT	-1.5±0.1	1.77±0.40	5.6±1.6	-7.25	-5.79
Mlh1 native ²⁰	pNtg2 (6-mère)	²³ RSKYFK	nd	nd	>180	nd	nd
Mlh1 native	pNtg2 (13-mère)	²⁰ VEVRSKYFKKNER	-20.5±1.5	1.11±0.08	9.0±0.6	-7.00	+13.5
Mlh1 native ²¹	pNtg2 (16-mère)	¹⁸ EEVEVRSKYFKKNERT	-26.1±4.6	1.40±0.3	7.1±2.4	-7.11	+19.0
MLH1 native ²²	pEXO1 (8-mère)	⁵⁰² TRSFFCS	-31.5±0.6	0.45±0.01	21.9±0.6	-6.42	+25.1
MLH1 native	pEXO1 (16-mère)	⁴⁹⁹ VVGTRSRFFCSSDST	-21.9±0.4	0.44±0.01	22.7±0.5	-6.40	+13.75

¹⁷ Les courbes expérimentales de BSI sont présentées en ANNEXE 7

¹⁸ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (J)

¹⁹ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (K)

²⁰ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (L)

²¹ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (M)

²² Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (N)

Influence de la température de mesure de l'interaction : Nous avons choisi d'évaluer les variations d'enthalpie de la réaction en fonction de la température afin de mesurer la variation de capacité calorifique, ΔC_p° , de l'interaction. Il a en effet été montré que la valeur et le signe de la capacité calorifique donne des informations quant à la nature des interactions observées (Stites, 1997). Pour ce faire, nous avons étudié l'interaction entre la protéine MLH1 et le peptide pNtg2(13-mère) à trois températures différentes 6°C, 18°C et 30°C (**Tableau 3 - 7**).

Tableau 3 - 7 : Analyses thermodynamiques par calorimétrie à titrage isotherme de l'interaction entre la région C-terminale de Mlh1 et du peptide de Ntg2 en fonction de la température et la concentration en sels

Protéine	Peptide	Température (°C)	[NaCl] (mM)	ΔH° (kcal/mol)	K_a ($10^5 M^{-1}$)	K_d (μM)	ΔG° (kcal/mol)	$-T\Delta S^\circ$ (kcal/mol)
Mlh1 native	pNtg2	6	300	-25.0±0.1	27.4±0.6	0.36±0.01	-8.19	+16.8
Mlh1 native	pNtg2	18	300	-26.9±0.1	8.4±0.1	1.19±0.02	-7.90	+19.1
Mlh1 native	pNtg2	30	300	-32.1±0.6	2.5±0.6	3.98±0.10	-7.46	+24.6
Mlh1 native ²³	pNtg2	6	100	-28.3±0.2	48.5±2.3	0.21±0.01	-8.51	+19.8
Mlh1 native ²⁴	pNtg2	6	300	-25.0±0.1	27.4±0.6	0.36±0.01	-8.19	+16.8
Mlh1 native ²⁵	pNtg2	6	600	-22.3±0.1	27.6±0.5	0.36±0.01	-8.20	+14.1
Mlh1 native ²⁶	pNtg2	30	100	-30.5±0.4	2.9±0.1	3.45±0.10	-7.55	+22.95
Mlh1 native	pNtg2	30	300	-32.1±0.6	2.5±0.6	3.98±0.10	-7.46	+24.6

L'analyse des thermogrammes montre que l'affinité augmente lorsque la température diminue. Les thermogrammes mesurés à 6°C présentent des profils permettant une détermination plus précise de l'enthalpie et de la constante de dissociation de la réaction. Nous pouvons en effet observer sur les thermogrammes et isothermes réalisés à 6°C, une meilleure définition des plateaux de pré- et de post-saturation. D'autre part, la variation d'enthalpie permet d'avoir une estimation de la variation de la capacité calorifique ΔC_p° , définie par $\Delta C_p^\circ = (\partial \Delta H / \partial T)_P$. En première approximation, la variation de la capacité calorifique ΔC_p° peut être estimée comme la pente de la droite reliant la variation enthalpique en fonction de la température. De cette façon, on obtient une variation de capacité calorifique ΔC_p° de -291 cal.deg⁻¹.mol⁻¹. Cette valeur correspond à une variation assez importante par rapport aux systèmes protéines-ligands reportés (Cole and Garsky, 2001). Les variations négatives de capacité calorifique sont typiquement associées à des interactions protéine-protéine mettant en jeu des interactions entre acides aminés hydrophobes. Elles peuvent également correspondre à des interactions mettant en jeu des changements conformationnels (Stites, 1997). Dans notre cas, il est probable que la valeur négative du facteur ΔC_p° correspond à l'enfouissement des deux aromatiques du motif MIP.

²³ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (F)

²⁴ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (G)

²⁵ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (H)

²⁶ Le thermogramme et l'isotherme de cette expérience sont visibles en ANNEXE 6 – (I)

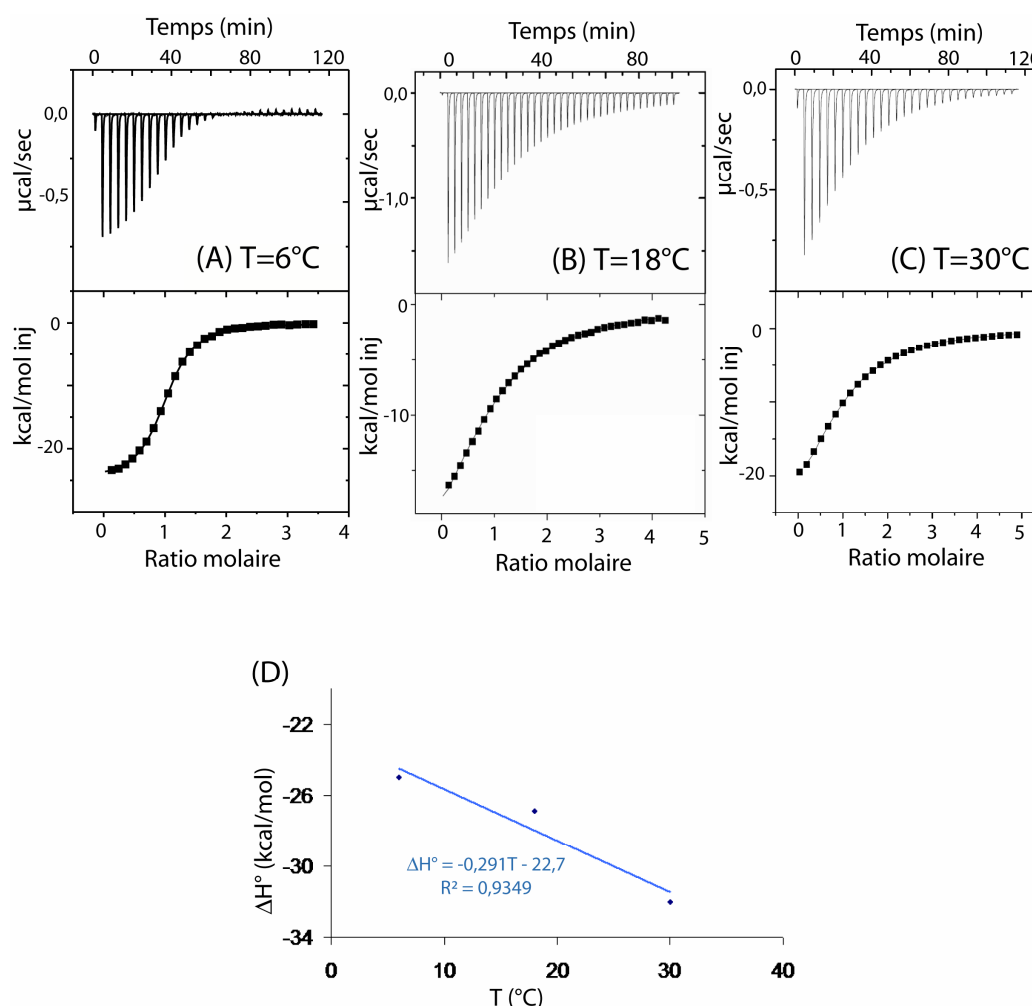


Figure 3 - 8 : Effet de la température sur les mesures par calorimétrie de l'interaction du peptide Ntg2 avec Mlh1 (A) Mesure à 6°C entre le domaine C-terminal de Mlh1 à $15\mu\text{M}$ et le peptide pNtg2(13-mère) à $100\mu\text{M}$ (B) Mesure à 18°C et le peptide pNtg2(13-mère) à $150\mu\text{M}$ (C) Mesure à 30°C et le peptide pNtg2(13-mère) à $150\mu\text{M}$ (D) Graphe de la variation enthalpique en fonction de la température ; la courbe de régression linéaire est indiquée en bleu clair

Influence de la concentration en sels dans le tampon de mesure: Enfin, nous avons décidé de varier la concentration en sels afin d'évaluer l'importance des interactions électrostatiques dans la formation du complexe entre Mlh1 et le peptide Ntg2 (**Tableau 3 - 7**). Nous avons réalisé les mesures à trois concentrations en NaCl : 100mM, 300mM et 600mM. La concentration de 300mM est la concentration standard utilisée dans les autres essais. L'augmentation de 300mM à 600mM en sels n'a d'influence ni sur l'affinité ni sur l'enthalpie et l'entropie. Le passage de 300 à 100mM en NaCl entraîne une légère amélioration de l'affinité d'un facteur 1.7 à 6°C ($K_d=0.21\mu\text{M}$ à 100mM en NaCl contre $K_d=0.36\mu\text{M}$ à 300m). Ce gain d'affinité est également observé à 30°C . Ces résultats suggèrent que les nombreux acides aminés, chargés positivement, de Ntg2 dans le motif MIP et à côté de ce motif (séquence RSKYFKK) contribuent légèrement à l'interaction.

En conclusion, les analyses par calorimétrie montrent que les acides aminés essentiels du motif MIP, (R/K)SK(Y/F)F, sont la serine et les deux aromatiques. L'interaction présente une affinité moyenne dans la gamme du micromolaire. Elle est résistante à la présence de sels. Les peptides d'une longueur de 8 acides aminés, c'est-à-dire un ou deux acides aminés supplémentaires de part et d'autre du motif, présentent la même affinité que les peptides de plus grande longueur.

3.2 ANALYSE STRUCTURALE DU DOMAINE C-TERMINAL DE *MUTL α*

3.2.1 ETUDE STRUCTURALE DU DOMAINE C-TERMINAL DE MLH1

i. Cristallogénèse du CTD de Mlh1 (485-769) de S. Cerevisiae

Dans le but de préciser le mode d'interaction entre le site S2 de Mlh1 et des peptides contenant un motif MIP, nous avons entrepris de cristalliser le domaine C-terminal de Mlh1 avec certains des peptides analysés en ITC.

Pour les tests de cristallogénèse, la protéine est préparée dans un tampon 25mM Na₂HPO₄ pH=7.5, 300mM NaCl, 10mM β -mercaptoéthanol et à une concentration comprise entre 3 et 12mg/ml. Les essais sont réalisés sur la protéine seule à 7.3mg/ml ou en complexe avec le peptide pNtg2(13mère) à deux concentrations différentes en protéine, 3.4mg/ml et 6.9mg/ml. Le mélange entre la protéine et le peptide est effectué pour avoir une stoechiométrie protéine:peptide de 1:3.

Le crible de conditions de cristallisation a été effectué sur la plateforme de cristallogénèse HTX du PSB (Partnership for Structural Biology) à Grenoble. Ce projet a fait l'objet d'une collaboration étroite avec Jose A. Marquez qui dirige la plateforme. Les projets ont été soumis et sélectionnés par la plateforme PCUBE (Protein Production Platform). Le crible effectué porte sur 576 conditions de cristallisation, correspondant aux kits Qiagen 1 à 6. La méthode utilisée est la cristallisation par diffusion de vapeur en goutte assise avec un mélange de 100nl de solution précipitante et 100nl d'échantillon, et ce, pour les trois échantillons précités (Mlh1(CTD) à 7.3mg/ml, Mlh1(CTD)*pNtg2 à 3.4mg/ml et Mlh1(CTD)*pNtg2 à 6.9mg/ml). Les plaques de cristallisation sont réalisées avec un robot Cartesian et sont analysées avec le robot de visualisation Formulatrix RockImager de la plateforme HTX située en chambre froide à 4°C. Des images des essais de cristallisation sont prises à différents temps (24h, 72h, 7 jours, 16 jours, 5 semaines et 10 semaines) et sont accessibles par internet depuis notre laboratoire. De ces 576 conditions, 46 ont été identifiées comme encourageantes, c'est-à-dire formant des microcristaux, des sphérulites ou du précipité hétérogène²⁷. Mais parmi ces 46

²⁷ Le détail des 46 conditions de cristallisation est donné en ANNEXE 8

conditions, une majorité comporte des cations métalliques. Or, la présence d'anions phosphate dans le tampon de la protéine rend probable la formation de cristaux de sels entre le phosphate et les cations métalliques utilisés dans les cribles. Nous avons sélectionné 9 conditions de cristallisation issues des résultats du robot pour les reproduire au laboratoire. Sur ces 9 conditions, nous avons obtenu des cristaux pour 5 d'entre elles. Des cristaux obtenus en présence de lithium(II), en forme de grains de riz se sont avérés être des cristaux de sels lors de tests de diffraction. De même, des sphérulites obtenues en présence de sulfate d'ammonium sont reproduits au laboratoire même en absence de protéine. A l'issue de ces essais, une seule condition retient l'attention; celle-ci contient du PEG 3.35k, agent précipitant le plus présent dans toutes les conditions retenues et de l'iodure de sodium à la concentration de 0.2M. Les microcristaux identifiés dans cette condition apparaissent seulement en présence du complexe Mlh1-peptide.

Parallèlement au crible effectué en robot, nous avons également testé d'autres cribles manuellement et ce, dans les mêmes conditions, à l'exception de la concentration en NaCl que nous avons choisi de réduire à 100mM. Cette diminution de concentration a pour but de réduire l'impact systématique de la présence de ce sel dans les essais de cristallisation. L'utilisation du crible proposé par E. Stura, appelé Footprint#1 (Molecular Dimensions Ltd) a permis d'identifier une seconde condition contenant 20% PEG 4k et 0.2M Imidazole Malate pH7. Cette dernière condition donne des aiguilles clairement identifiables. Comme pour la première condition, ces cristaux n'apparaissent qu'en présence du complexe Mlh1(CTD)*peptide pNtg2.

Nous avons alors cherché à optimiser, de façon conjointe, les deux conditions (PEG 3.35k/NaI et PEG4k/Imidazole Malate). Les deux conditions ont été identifiées en présence du peptide pNtg2. Nous avons d'abord testé les différents peptides pExo1 et pSgs1 et avons observé la formation de cristaux mais nettement plus petits. Nous avons décidé de tester également un certain nombre de paramètres: concentration en agent précipitant, concentration en complexe, pH, présence d'additifs, présence d'agent réducteur et stœchiométrie du complexe protéine/peptide, méthode de cristallisation (diffusion de vapeur, en batch sous huile), la température de cristallisation (4°C ou 20°C). Le seul paramètre qui a permis une réelle amélioration des cristaux obtenus avec la condition (PEG 3.35k/ NaI) a été la cristallisation en batch sous huile. Les meilleurs cristaux obtenus ont une forme de bâtonnets et une taille inférieure à 5µm dans deux dimensions (**Figure 3 - 9(A)**). Ces cristaux ont été congelés et analysés sur les lignes synchrotron ID23-2 (ESRF-Grenoble). Ils ne donnent aucune tâche de diffraction.

La seconde condition (PEG 4k/Imidazole Malate) a subi plusieurs étapes d'optimisation: d'une part, l'ajout de l'agent réducteur Dithiothreitol (DTT) à une concentration comprise entre 20 et 100mM permet d'obtenir des aiguilles plus grandes d'un facteur 2 dans toutes les dimensions; d'autre part, la cristallisation en batch

sous huile de paraffine plutôt qu'en diffusion de vapeur permet d'obtenir des aiguilles de tailles acceptables pour les essais de diffraction à savoir $200 \times 10 \times 5 \mu\text{m}^3$. Néanmoins dans ces conditions, beaucoup de ces cristaux forment un tapis en appui sur le plastique de la boîte de cristallisation. Dans ces conditions, les cristaux sont difficiles à récupérer et à congeler.

Afin d'éviter une très forte nucléation, nous avons réalisé des gouttes en utilisant l'agent précipitant, le PEG 4k, à une concentration plus faible, inférieure à 10%(w/v). A cette concentration de PEG, la nucléation n'est pas spontanée. Une fois la goutte équilibrée, quelques microcristaux préalablement lavés et broyés sont ajoutés à la goutte. Nous avons alors observé que les microcristaux introduits servent de support à la croissance d'aiguilles qui forment des oursins, à la base desquels se trouve le microcristal. Cet ensemencement permet d'obtenir peu de cristaux dans chaque goutte et ceux-ci sont d'une part, plus faciles à congeler et d'autre part, plus épais, avec une épaisseur et une largeur de $10 \mu\text{m}$ (**Figure 3 - 9(B)**).

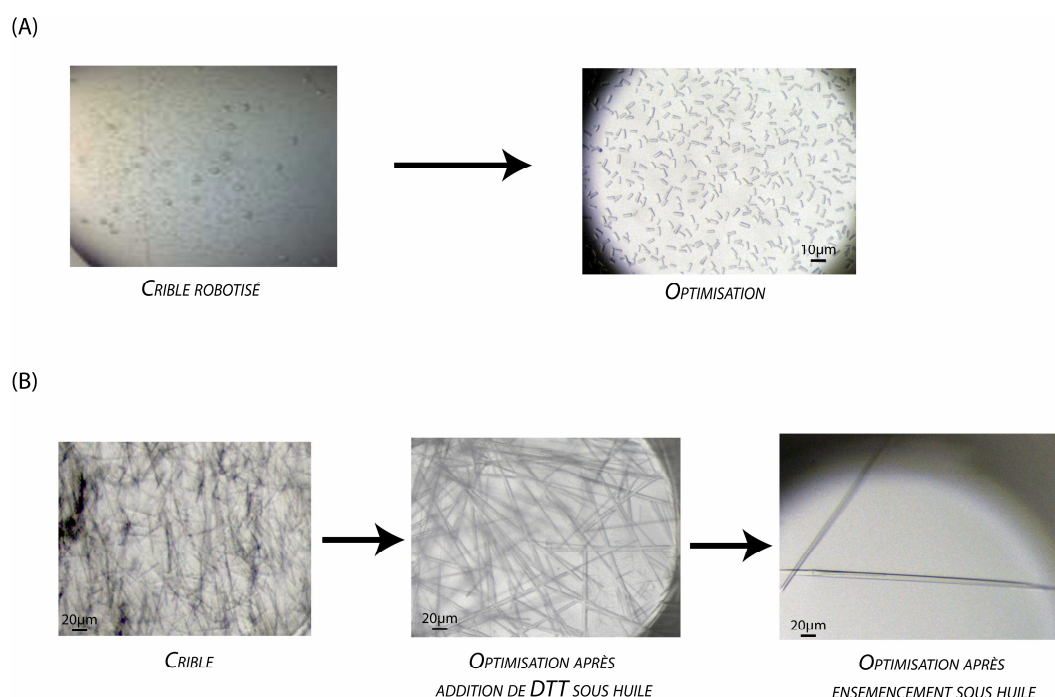


Figure 3 - 9: Cristallogénèse du complexe entre le domaine C-terminal de Mlh1 et le peptide Ntg2 (A) (Gauche) Sphérulites obtenues en présence de 20% PEG 3.35k et 0.2M NaI sur la plateforme HTX. (Droite) Meilleurs cristaux obtenus au laboratoire autour de cette condition (28% PEG 3.35k; 0.4M NaI) sous huile de paraffine; Le complexe Mlh1/pNtg2 est à une concentration de 8mg/ml en Mlh1 (B) (Gauche) Aiguilles obtenues au laboratoire à partir du cribble Footprint #1 (20% PEG 4k; 0.2M Imidazole Malate pH7). (Centre et droite) Etapes d'optimisation en présence de 22% PEG 4k, 0,4M Imidazole Malate pH7; 20mM DTT sous huile de paraffine; l'échantillon de complexe Mlh1/pNtg2 est à une concentration de 8mg/ml en Mlh1

Ces aiguilles, une fois congelées dans 20% glycérol, ont été analysées sur les lignes synchrotron Proxima1 (Soleil, Gif-sur-Yvette). Ces cristaux possèdent un faible pouvoir de diffraction et se dégradent très rapidement (moins de 50 images de

rotation 1°). Seules quelques images avec des tâches de diffraction à 7Å ont pu être obtenues.

ii. Cristallogenèse du CTD de Mlh1 (503-769) de S. cerevisiae

Devant les premières difficultés rencontrées pour obtenir des cristaux de la région C-terminale de Mlh1 seul ou en complexe avec un peptide de pouvoir diffractant satisfaisant, nous avons entrepris de purifier une forme plus courte du CTD de Mlh1. En effet, les premiers résidus de l'extrémité N-terminale sont prédits comme non structurés. Ces résidus semblent faire partie de la longue région prédite comme non-structurée reliant les domaines N- et C-terminaux de Mlh1 (acides aminés 335 à 507 environ). Afin d'ôter les extrémités non-structurées éventuellement présentes en N-terminal ou C-terminal de la région (485-769) de Mlh1 utilisée dans nos premiers essais de cristallisation, nous avons effectué des tests de protéolyse ménagée à la trypsine. Nous avons incubé pendant différents temps 100µl du CTD de Mlh1 à 1mg/ml avec 0,1% m/m de trypsine (1µl de trypsine à 100µg/ml) dans le tampon Tris à pH8. Nous observons sur gel SDS une forme plus courte de quelques centaines de daltons par rapport à la protéine initiale (poids moléculaire initial environ 33kDa) qui apparaît dès 15min d'incubation. Cette forme tronquée est majoritaire dès 15min d'incubation, à température ambiante ou à 37°C et devient la seule forme observable sur gel SDS au bout de 3h à 37°C (**Figure 3 - 10(A)**).

Le fragment protéolysé a ensuite été purifié sur colonne d'exclusion de taille (SEC pour Size Exclusion Chromatography en anglais) puis analysé par spectrométrie de masse (MALDI-TOF) et séquençage N-terminal. Le spectre de masse montre quatre pics majoritaires à 30763, 30479, 30235 et 29948Da par rapport à un poids attendu pour le CTD de Mlh1 (485-769) non-protéolysé de 33198Da de la protéine (**Figure 3 - 10(B)**). Le séquençage N-terminal de l'échantillon protéolysé en solution montre deux séquences: une majoritaire, ₅₀₅ERVNVNLTSL et une minoritaire, ₅₀₇VNVNLTSLKK. Les formes majoritaires et minoritaires correspondantes aux résultats du séquençage N-terminal ont un poids moléculaire théorique de 30727 et 30442Da respectivement dans l'hypothèse où aucune protéolyse n'a eu lieu à l'extrémité C-terminale. Ces masses sont en bon accord avec les valeurs obtenues par spectrométrie de masse.

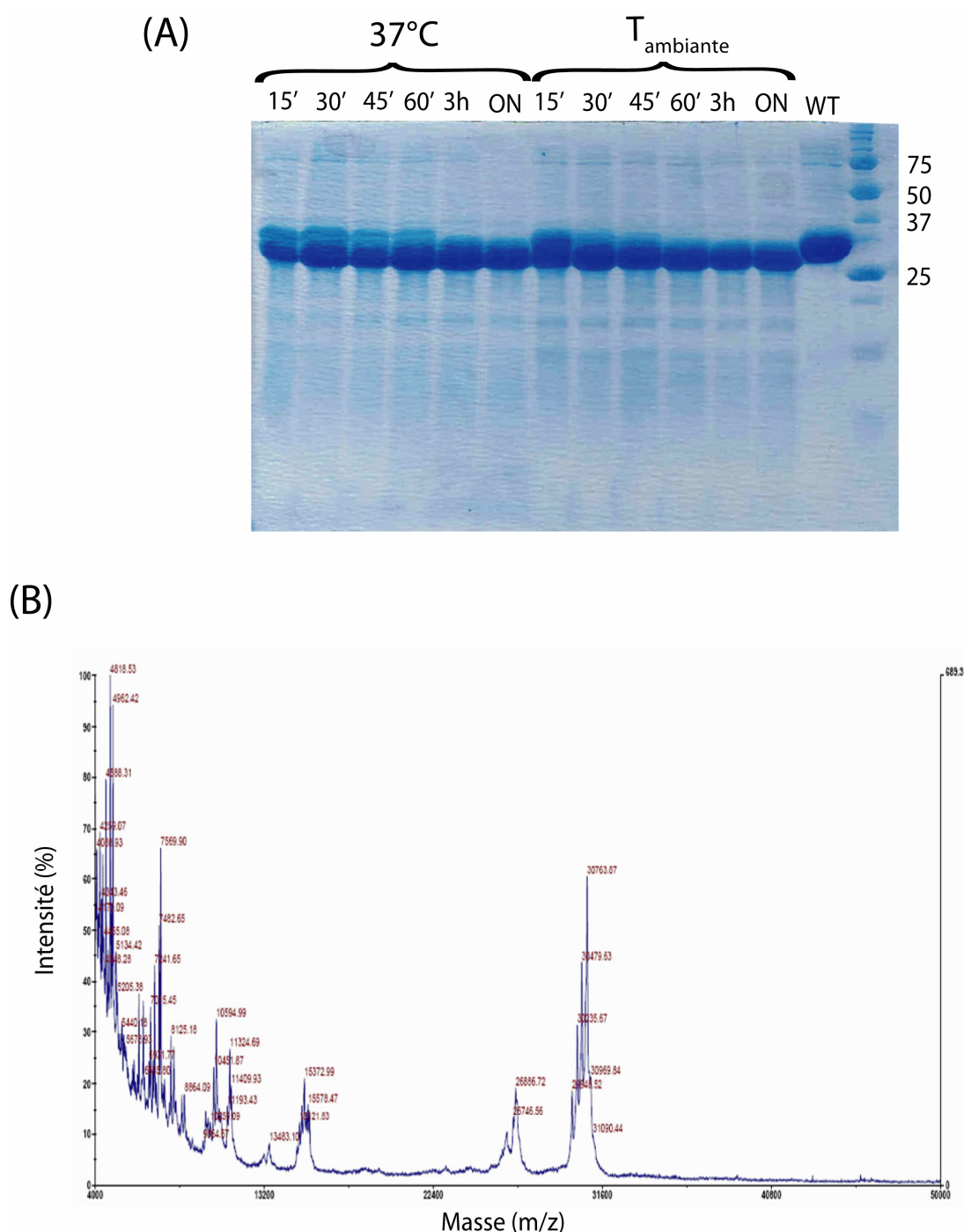


Figure 3 - 10: Protéolyse ménagée à la trypsin, du domaine C-terminal de Mlh1 (A) Analyses sur gel SDS en conditions dénaturantes de la trypsinolyse du CTD de Mlh1 en fonction de la température: 100µl du CTD de Mlh1 à 1mg/ml est mélangé à 1µl de trypsin à 100µg/ml et 10µl Tris pH8 à 1M (B) Spectre de masse du fragment protéolysé après passage sur colonne d'exclusion stérique

Sur la base de ces résultats, une seconde construction (503-769) de Mlh1(CTD), nommée Mlh1(CTD-v2) a été clonée. Cette construction a été sur-exprimée et purifiée selon le même protocole que Mlh1(CTD-v1) (485-769) précédemment étudiée. Nous avons évalué par calorimétrie l'interaction entre Mlh1(CTD-v2) et le peptide pNtg2 par ITC. Le thermogramme et l'isotherme d'interaction sont en tout point comparable à ceux obtenus avec Mlh1(CTD-v1) (données non montrées). Un second

crible a été réalisé sur la plateforme HTX avec cette nouvelle construction Mlh1 (503-769). Les 576 conditions de cristallisation ont été testées avec le complexe Mlh1: peptide pNtg2(13-mère) dans un rapport 1:3. Ce crible a été réalisé à trois concentrations différentes en complexe, *i.e.* 3, 6 et 12mg/ml et en utilisant le tampon 20mM HEPES, 150mM NaCl afin d'éviter la formation de cristaux de sels obtenus précédemment avec le tampon Na₂HPO₄. Des cristaux ont été observés dans 30 conditions parmi les 576 testées et dans 29 autres conditions, nous avons observé des microcristaux, des sphérulites ou du précipité hétérogène²⁸. Les conditions de cristallisation utilisant des cations sont beaucoup moins représentées que dans le crible précédent où le complexe était en tampon phosphate. Parmi les 30 premières conditions, quasiment toutes présentent des aiguilles en présence de PEG, exception faite pour deux conditions ayant comme agent précipitant l'isopropanol. Nous avons sélectionné 10 conditions de cristallisation issues des résultats du robot pour les reproduire au laboratoire. Sur ces 10 conditions, nous avons obtenu des cristaux pour 7 d'entre elles. Comme nous l'avons également constaté pour la construction Mlh1 (485-769), la construction Mlh1(CTD-v2) en complexe avec le peptide pNtg2 cristallise majoritairement en présence de PEG 3.35k ou PEG 4k. Reproduites au laboratoire, ces conditions ont permis d'obtenir des aiguilles en tous points similaires à celles obtenues précédemment. Nous avons notamment observé de façon semblable une amélioration de la cristallisation avec l'ajout de DTT et la réalisation des expériences de cristallisation sous huile de paraffine. Les cristaux obtenus ont été testés sur la ligne Proxima1 du Synchrotron Soleil (Gif-sur-Yvette). Ils ont un pouvoir de diffraction très faible, *i.e.* vers 8Å.

En conclusion, les essais de cristallisation sur la région C-terminale de Mlh1 seul ou en complexe avec un peptide contenant le motif MIP n'ont pas permis d'obtenir des cristaux de pouvoir diffractant satisfaisant malgré un nombre d'essais de cristallisation important (environ 1200 essais au robot et 2000 essais réalisés manuellement au laboratoire).

A l'issue de ces premiers travaux, nous avons choisi de rediriger nos efforts sur la cristallisation de la région CTD de Mlh1 en complexe avec la région CTD de Pms1 et donc de purifier la région C-terminale de MutL α . La protéine Mlh1 est en effet présente dans la cellule, essentiellement en complexe avec d'autres homologues de MutL, Pms1, Mlh2 et Mlh3. Comme cela a été évoqué en introduction, il a été proposé que la région principale d'hétérodimérisation de Mlh1 soit la région C-terminale. Le complexe Mlh1(CTD)*Pms1(CTD) correspond à une conformation de la région C-terminale de Mlh1 plus proche celle observée dans la cellule. Ce complexe est donc susceptible de présenter un comportement plus favorable d'un point de vue

²⁸ Le détail des 59 conditions de cristallisation est donné en ANNEXE 9

biochimique que la région Mlh1(CTD) seule ou en complexe avec un peptide. La seconde partie de ma thèse a porté sur l'étude ce complexe CTD Mlh1*CTD Pms1.

3.2.2 RESOLUTION DE LA STRUCTURE CRISTALLOGRAPHIQUE DE LE REGION C-TERMINALE DU COMPLEXE MLH1*PMS1 DE *S.*

CEREVISIAE

*i. Purification et caractérisation biophysique du complexe de Mlh1(CTD)*Pms1(CTD)*

Afin d'étudier le complexe CTD de Mlh1 / CTD de Pms1, nous avons choisi de reprendre pour Mlh1 les deux constructions utilisées précédemment qui permettent de purifier la protéine à homogénéité et en quantité suffisante pour des études structurales. Pour le CTD de Pms1, les cribles de double-hybride montrent une région minimale d'interaction pour Pms1 de levure de 635 à 873 (Kondo et al., 2001). Différentes approches sont envisageables pour la production d'un complexe protéine-protéine. On peut envisager la purification des protéines séparément, leur co-purification à partir de lysat des protéines exprimées séparément ou la co-expression et co-purification de ces protéines. Des essais antérieurs de purification du CTD de Pms1, menés au laboratoire avant mon arrivée en thèse par C. Quemener, avaient montré que le CTD de Pms1 en fusion avec un tag 6×His ou en fusion avec l'étiquette MBP est très peu soluble. Nous avons centré nos efforts sur la co-expression et la co-purification du complexe comprenant le domaine C-terminal de MutL α de *S. cerevisiae*, *i.e.* les domaines 485-769 de Mlh1 et 635-873 de Pms1²⁹.

Les domaines C-terminaux de Mlh1 et de Pms1 sont clonés sur des vecteurs séparés portant des résistances kanamycine et ampiciline respectivement et des origines de réplifications différentes. Les protéines sont co-exprimées après transformation des cellules compétentes contenant le plasmide codant pour GST-Mlh1(CTD) par le plasmide codant MBP-Pms1(CTD). Mlh1(CTD) est cloné en fusion avec l'étiquette 6×His-GST et Pms1(CTD) en fusion avec l'étiquette 6×His-MBP. Les fusions 6His-MBP-Pms1 et 6His-GST-Mlh1 possèdent un site de clivage à la protéase TEV, entre l'étiquette et le domaine C-terminal. Les extraits bruts, avant et après induction des cultures d'expression du complexe, sont analysés sur gel SDS-PAGE. Le gel indique une forte surexpression d'une bande à la hauteur de la protéine de fusion MBP-Pms1 (PM 70kDa) et d'une bande à la hauteur de la protéine de fusion GST-Mlh1 (PM 62 kDa) (résultat non montré). Le surnageant, après lyse cellulaire et centrifugation de 1L de culture de *E. coli* avec le CTD de MutL α , est chargé sur une résine d'amylose puis élué par le maltose, ligand naturel de l'étiquette MBP (**Figure 3 - 11(A)**). A l'issue de cette étape, environ 50mg du complexe MBP-

²⁹ Le détail des protocoles de purification est disponible en ANNEXE 10

Pms1(CTD)*GST-Mlh1(CTD) sont obtenus par litre de culture d'*E. coli* (les quantités de protéine sont estimées par la méthode de Bradford).

L'éluat de la colonne d'amylose est coupé avec 2% m/m (masse/masse) de protéase TEV à 4°C sur la nuit avec agitation douce. Les protéines portant une étiquette 6×His sont éliminées par passage sur une colonne de nickel. Le complexe d'intérêt passe directement et se retrouve dans le filtrat direct (FT ou flow through en anglais) (**Figure 3 - 11(A)**). En l'absence complète d'imidazole dans le tampon de l'échantillon élué d'amylose, le complexe est retenu de façon non-spécifique sur la colonne et s'élue à la fin du chargement, lorsque la colonne est lavée avec le tampon d'équilibration contenant 20mM d'imidazole. Dans les purifications suivantes, 20mM d'imidazole sont ajoutés à l'échantillon, selon les recommandations du fournisseur, et chargés sur la colonne pour éviter cette fixation non-spécifique. Le rendement de cette étape est de l'ordre de 50% puisque l'on obtient environ 12mg de CTD de MutL α par litre de culture. La quantité maximale de complexe attendue pour 50mg de fusion éluée est en effet de 25mg, les 25mg restants correspondant à la GST et à la MBP comptabilisées dans la fraction éluée d'amylose (concentrations estimées par la méthode de Bradford). L'analyse sur gel SDS des fractions issues de la coupure à la TEV et de la colonne de nickel suggère que la perte d'une partie du complexe CTD de MutL α à cette étape provient d'une précipitation d'une fraction de Pms1 lors de la coupure à la TEV.

Le CTD de MutL α est ensuite dialysé contre un tampon 20mM Na₂HPO₄ pH7.4 contenant 50mM NaCl et est passé sur une colonne échangeuse d'anions de type Resource Q (**Figure 3 - 11(B)**). Le chromatogramme obtenu présente un pic à 280nm pendant le gradient en NaCl. Ce pic sort à une conductivité de 20mS.cm⁻¹ et correspond, d'après les gels SDS, au CTD de MutL α . Le second pic, très faible, sort à une conductivité de 50mS.cm⁻¹ et possède une très forte absorption à 254nm. L'analyse sur gel SDS suggère que ce pic correspond à des fragments d'ADN. L'analyse des spectres UV du CTD de MutL α sous forme de protéines de fusion ou après passage sur la colonne de nickel indique que le complexe fixe des fragments d'ADN générés lors de la lyse cellulaire. Une grande partie des fragments d'ADN fixés est éliminée par une étape de dialyse. L'ADN encore présent est éliminé lors de l'étape de polissage sur Resource Q. La pureté de l'échantillon à l'issue de cette étape de polissage est compatible avec des essais de cristallisation. Le rendement de cette dernière étape sur résine échangeuse d'anions est faible, de l'ordre de 20%. Lors de certaines purifications, nous avons également observé que le complexe obtenu ne présente plus une stoechiométrie Mlh1(CTD): Pms1(CTD) de 1:1. La sous-unité Mlh1 est présente en quantité plus importante que celle de la sous-unité Pms1, suggérant qu'une fraction de la protéine Pms1 précipite au cours de cette étape ou reste agrégée sur la colonne. Afin d'optimiser cette étape de polissage, nous avons essayé de

remplacer l'étape d'échange d'anions par une étape de chromatographie "d'interactions hydrophobes" sur une colonne de type Resource Phe (**Figure 3 - 11(C)**). On ajoute alors 1.5M $(\text{NH}_4)_2\text{SO}_4$ à l'échantillon avant son passage sur la colonne. A l'issue de cette étape, le complexe Mlh1*Pms1 présente une stoechiométrie plus proche de la valeur 1 :1 recherchée mais le rendement n'est pas meilleur (environ 20%). Le rendement faible observé dans le cas de la colonne d'interaction hydrophobe semble provenir d'une précipitation du complexe au départ lorsqu'il est mis en contact avec une très forte force ionique pour s'accrocher à la colonne.

Au final, les meilleures purifications ont permis d'obtenir environ 3mg de la région C-terminale du complexe Mlh1-Pms1 pour 1 litre de culture d'*E. coli*. La purification de ce complexe reste délicate car pour des raisons que nous ne comprenons pas encore complètement, les deux protocoles de purification avec polissage par échangeuse d'anions ou par interaction hydrophobe ne permettent pas d'obtenir systématiquement des échantillons du CTD de MutL α en quantité suffisante ($m > 1\text{mg}$), avec une bonne stoechiométrie (proche de 1:1) et une pureté satisfaisante (somme des bandes Mlh1 et Pms1 sur gels SDS représentant 90% des protéines totales). Il est encore fréquent qu'à l'issue d'une purification un ou plusieurs de ces critères ne soient pas satisfaits. Cette observation suggère que ces protocoles méritent un travail d'optimisation supplémentaire afin d'obtenir de façon reproductible des échantillons en quantité et qualité compatible avec des études structurales.

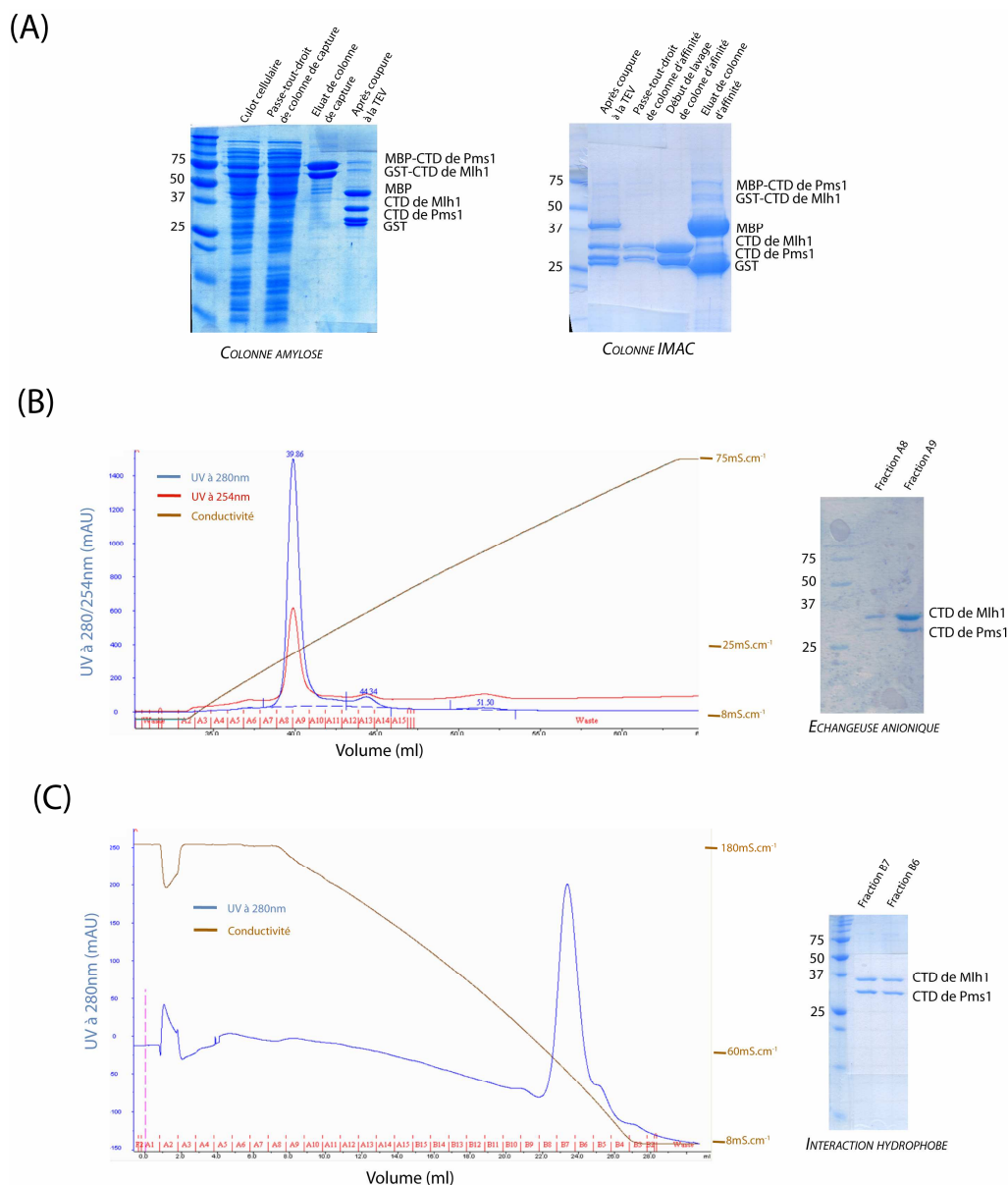


Figure 3 - 11: Exemple de gels SDS et chromatogrammes obtenus lors des étapes de purification du domaine C-terminal de Mlh1*Pms1 (A) Gels SDS de fractions issues des deux premières étapes de purification du CTD de Mlh1*Pms1 (B) Profil d'élution de la colonne échangeuse d'anions et gels SDS du CTD de Mlh1*Pms1 (C) Profil d'élution de la colonne d'interaction hydrophobe et gels SDS du CTD de Mlh1*Pms1

ii. Cristallo-genèse du domaine C-terminal de Mlh1*Pms1 et collecte des données natives

Avant de lancer les tests de cristallisation, l'échantillon est dialysé contre 25mM Tris pH=7.5, 300mM NaCl, 1mM TCEP, 1mM AEBSF et est concentré entre 4 et 10mg/ml. Un crible de conditions de cristallisation est effectué sur la plateforme de cristallisation HTX (PSB-Grenoble) selon un protocole semblable à celui adopté pour le complexe Mlh1(CTD)*peptide de Ntg2. Brièvement, 576 conditions de cristallisation sont testées en diffusion de vapeur en mélangeant 100nl de solution précipitante avec 100nl de l'échantillon, et ce, pour deux concentrations du complexe

(4.4mg/ml et 7mg/ml). Les boîtes sont analysées à 4°C par un robot de visualisation. De ces 576 conditions, 42 présentent des cristaux clairement définis et notamment de nombreuses aiguilles (**Figure 3 - 12(A)**). 10 autres conditions sont également annotées car elles représentent des pistes potentiellement intéressantes (microcristaux, sphérulites ou des précipités cristallins)³⁰.

Sur les 42 conditions les plus nettes, 20 conditions ont comme agent précipitant des PEG entre 3.35k et 8k (6, le PEG 4K; 6, le MPEG 5K; 5, le PEG 8K, 3 le PEG 3.35K). 16 conditions ont comme agent précipitant un solvant organique (11, le 2-Méthyl-2.4-pentanediol (MPD), 5, l'isopropanol). Enfin 3 conditions ont comme agent précipitant un PEG de faible poids moléculaire, le PEG 400. On peut noter que tous ces agents précipitants sont des composés chimiques présentant des fonctions hydroxyles que ce soit directement des alcools ou des polymères d'éthylène glycol (PEG). Le pH de ces 42 pistes de cristallisation est compris entre 6 et 9, c'est-à-dire au-dessus du point isoélectrique théorique du complexe (pI=5.4). Enfin, il est intéressant de noter que dans de nombreuses conditions, la solution de cristallisation contient du magnésium.

Du point de vue de la taille des cristaux et de la régularité de leur forme, les meilleurs agents précipitants identifiés, à l'issue du crible effectué sur le robot de la plateforme HTX, sont le MPD, l'isopropanol, le PEG 6K et le MPEG 5K. Chacun donne des aiguilles fines (environ 1µm de largeur et d'épaisseur) et relativement longues (de l'ordre de 100µm). Ces aiguilles croissent rapidement puisqu'elles apparaissent dès le lendemain du jour où a été réalisé le crible sur la plateforme HTX. A partir de ces résultats, nous avons choisi 4 conditions pour les reproduire au laboratoire avec des gouttes de 3µl au lieu de 0.2µl. Ces conditions ont été testées avec des gammes de concentration en ces quatre agents précipitants encadrant la valeur observée avec le robot de cristallisation. Nous avons analysé également l'influence du pH avec les tampons suivants: 0.1M MES à pH6; 0.1M HEPES à pH7 et 0.1M Tris à pH8. Des essais ont été réalisés à 4°C mais également à 20°C. Pour cela, des essais ont été réalisés en testant des concentrations en agents précipitants encadrant la valeur observée avec le robot de cristallisation. Sur les 4 conditions sélectionnées pour être reproduites, toutes ont permis d'obtenir des cristaux au laboratoire. De nombreuses aiguilles sont obtenues de tailles proches de celles observées sur la plateforme HTX. Les concentrations en agent précipitant identifiées au laboratoire pour reproduire ces cristaux sont systématiquement plus faibles que celles trouvées par le robot. A cette étape, les plus beaux cristaux sont obtenus dans la condition suivante : 10% m/v (masse/volume) MPEG 5K, 0.1M HEPES pH7. Les aiguilles obtenues ont pour dimensions 100×3×3µm³. Ces aiguilles sont 9 fois plus volumineuses que celles obtenues au robot qui présentaient une largeur et une épaisseur de 1µm. Ces aiguilles ont ensuite été congelées dans 20% glycérol et ont

³⁰ Le détail des 52 conditions de cristallisation est donné en ANNEXE 11

permis d'avoir un premier jeu de données à 3.5Å, complet à 50% sur la ligne ID23-1 de l'ESRF (Grenoble).

Nous avons alors entrepris d'améliorer ces cristaux. Nous avons choisi de travailler sur les paramètres suivants : pH, ajout de sel, ajout d'additifs, méthode de cristallisation, ratio agent précipitant/protéine, concentration en agent précipitant ou en protéine, température, ensemencement. Parmi ces paramètres, deux ont eu un rôle important dans l'obtention de cristaux de taille supérieure : l'ajout en tant qu'additif de cations divalents et l'ensemencement.

En premier lieu, lors d'un crible en additifs, nous avons pu observer que l'ajout de cations Mg^{2+} , à une concentration de 10mM, favorisait la croissance des cristaux. Nous avons ainsi obtenu des cristaux significativement plus larges et plus épais, atteignant $200 \times 20 \times 10 \mu m^3$. Différents contre-ions associés au cation Mg^{2+} ont été testés. L'utilisation de chlorure, sulfate ou acétate de magnésium donne des cristaux de taille similaire. L'effet des ions magnésium est en accord avec le rôle de ce cation observé lors des cribles par le robot de cristallisation. De plus, l'influence de ce cation est à rapprocher des données de la littérature sur Pms1. Il est en effet proposé que le site endonucléase de Pms1 possède un site de fixation pour un ou plusieurs de ces cations Mg^{2+} , Mn^{2+} , Zn^{2+} ou Fe^{2+} (Kosinski et al., 2008). Les tests *in vitro* du laboratoire de P. Modrich suggèrent que l'activité de Pms1 nécessite principalement du Mg^{2+} . Plus récemment, une étude du laboratoire a proposé que Pms1 fixe au moins un zinc. Les essais en présence de manganèse, ajouté à une concentration comprise entre 1mM et 10mM, permettent d'obtenir de petites aiguilles, de qualité inférieure à celles obtenues en présence de magnésium. Les essais en présence d'ion Zn^{2+} et Fe^{2+} , dans la même gamme de concentration 1mM-10mM, font précipiter le complexe protéique, même à faible concentration en agent précipitant.

Dans un second temps, l'ensemencement de microcristaux a permis de mieux contrôler la croissance des aiguilles (**Figure 3 - 12(B)**). En effet, dans de nombreuses gouttes, le taux de nucléation, très fort, aboutit à un tapis d'aiguilles relativement fines. Afin d'inhiber la nucléation et de mieux contrôler la croissance de ces aiguilles, nous avons réalisé des gouttes avec une concentration en agent précipitant comprise entre 2 et 6%(m/v). Cette concentration permet de se placer dans une zone où la nucléation n'est pas spontanée. Une fois la goutte équilibrée par diffusion de vapeur, quelques microcristaux préalablement lavés et broyés sont ajoutés à la goutte. Nous avons observé comme pour la cristallisation du complexe Mlh1/peptide que les microcristaux introduits ne grossissent pas mais servent de support à la croissance des aiguilles. On obtient également des structures en forme d'oursin à la base de laquelle se trouve le microcristal. Il a été proposé dans une étude méthodologique de cristallogénèse (Thakur et al., 2007) que l'ajout de billes d'hydroxyapatite peut également jouer un rôle de catalyseur de nucléation de cristaux. Dans notre cas, cette approche a permis également de former de façon plus contrôlée et plus régulière des aiguilles de taille intéressante.

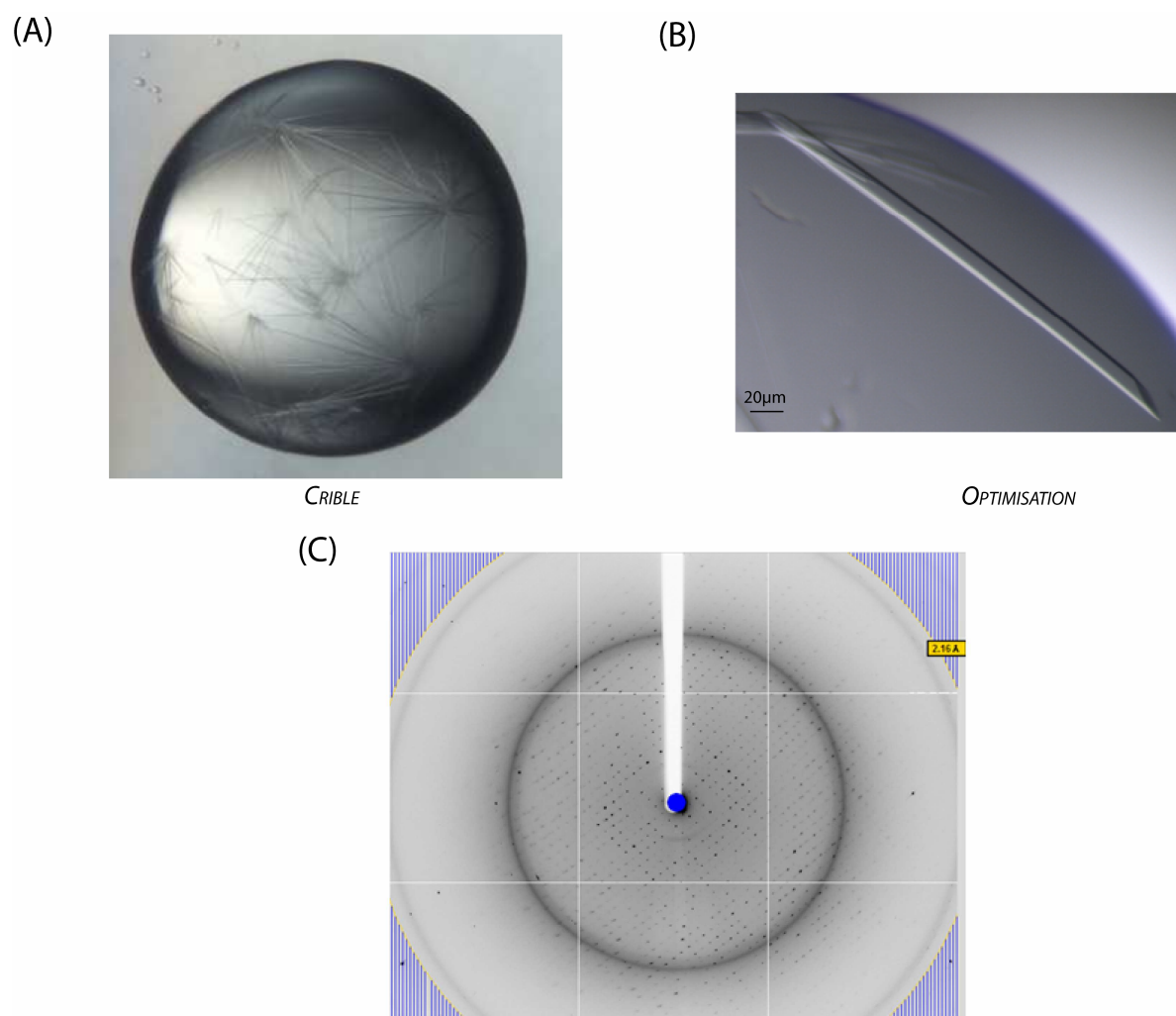


Figure 3 - 12 : Cristallisation du domaine C-terminal de MutL α (A) Aiguilles issues du crible réalisé sur la plateforme HTX ; la solution du réservoir contient 5% (m/v) MPEG 5K, 0.1M HEPES pH7 (Screen Mme 5000-Hampton) ; la solution du complexe protéique a une concentration de 4.8mg/ml (concentration mesurée par spectre UV) (B) Cristaux obtenus après optimisation au laboratoire; la solution du réservoir contient 6%(m/v) MPEG 5K ; 0.1M HEPES pH7 ; 10mM MgCl₂ ; la solution du complexe protéique a une concentration de 5mg/ml (C) Cliché de diffraction d'une aiguille obtenu sur la ligne de lumière Proxima 1 (Synchrotron Soleil) après 1s d'exposition, sans atténuation du faisceau, pour une rotation de 1°. Ces aiguilles diffractent à 2.6Å de résolution.

Cette stratégie a permis d'obtenir des cristaux de taille 200×20×10µm³, ayant un pouvoir de diffraction maximum de 2.6Å. Cependant, en relation avec leur morphologie, beaucoup de cristaux testés sur les lignes de lumière présentent une forte anisotropie: il n'est pas rare de voir des cristaux qui diffractent à 2.6Å selon une direction et 4Å selon une direction perpendiculaire. Ces cristaux sont également sensibles aux rayons X et perdent rapidement leur pouvoir de diffraction du fait de leur faible largeur et épaisseur. Par conséquent, un grand nombre de cristaux a dû être testé afin d'obtenir un jeu de données complet avec un minimum de redondance et à une résolution meilleure que 3.5Å.

Au final, les meilleurs cristaux sont obtenus à 20°C avec une solution de réservoir composée de 6%(m/v) MPEG 5K, 0.1M HEPES pH7, 10mM MgCl₂ puis sont congelés dans 30% glycérol. Les données de diffraction ont été enregistrées sur la

ligne de lumière Proxima1 (Synchrotron Soleil). Le meilleur cristal diffracte jusqu'à 2.6Å et appartient au groupe d'espace C2 avec les paramètres de maille suivants: $a = 187.1\text{\AA}$, $b = 66.1\text{\AA}$, $c = 74.0\text{\AA}$ et $\beta = 90.8^\circ$, $\alpha=\gamma=90^\circ$ (Figure 3 - 12(C)). Les statistiques des données de diffraction sont présentées dans le Tableau 3 - 8.

Tableau 3 - 8 : Statistiques des données de diffraction ; les valeurs entre parenthèses correspondent aux statistiques dans la dernière coquille de résolution

	Native
Collection des données	PX1
Groupe d'espace	C2
Paramètres de maille	
a, b, c (Å)	187.1, 66.1, 74.0
α, β, γ (°)	90.0, 90.8, 90.0
Longueur d'onde (Å)	0.9800
Résolution (Å)	34.54-2.60 (2.74-2.60)
R_{sym} or R_{merge}	9.70 (71.2)
$I / \sigma I$	10.5 (2.10)
Complétude (%)	99.4 (97.5)
Redondance	4.0 (4.0)

iii. Essais de phasage par remplacement moléculaire et par la méthode des dérivés d'atomes lourds

Essais de phasage par remplacement moléculaire

Le remplacement moléculaire prend une place de plus en plus importante comme méthode de phasage. Ceci est lié au nombre croissant de structures disponibles mais également au progrès des méthodes de modélisation moléculaire qui permettent dans de nombreux cas de créer un modèle assez ressemblant de la protéine ou du complexe étudié. Il est généralement admis que le remplacement moléculaire est une technique assez efficace pour obtenir des phases lorsque la ou les protéines étudiées présentent une identité de séquence d'au moins 30%. Dans notre cas, la structure du domaine C-terminal de MutL d'*E. coli*, ecMutL, est disponible. Plusieurs arguments suggèrent un repliement semblable entre les régions C-terminal de l'homodimère ecMutL et l'hétérodimère MutL α . (i) Les protéines ecMutL et MutL α présentent une similarité fonctionnelle importante au sein de la voie MMR. (ii) Les régions C-terminales constituent les régions d'homodimérisation et d'hétérodimérisation respectivement de ecMutL et MutL α . (iii) Les structures des régions N-terminales de ecMutL, hsMLH1 et hsPMS2 ont été résolues et présentent un repliement similaire. Cependant, le domaine C-terminal d'ecMutL ne possède que 9% et 12% d'identité de séquences avec, respectivement, le CTD de Mlh1 et Pms1 de levure. Les pourcentages de similarité de séquences sont également faibles avec 20%

de similarité entre ecMutL et les CTD de Mlh1 ou Pms1. Ces valeurs suggèrent des différences potentiellement importantes de structure 3D. En conclusion, le remplacement moléculaire semble une méthode intéressante en raison des repliements similaires attendus avec ecMutL mais difficile en raison du faible pourcentage d'identité de séquences.

Dans une première étape, nous avons constitué un ensemble de modèles susceptibles d'être utilisés en remplacement moléculaire. Ces modèles correspondent soit à des structures cristallographiques ou aux résultats de modélisation moléculaire par des approches différentes (**Figure 3 - 13(A)**). Nous avons ainsi utilisé:

- (i) La structure de l'homodimère du CTD de ecMutL (code PDB:1X9Z) (Guarne et al., 2004);
- (ii) Le modèle décrit dans la première partie des Résultats (§3.1.1), construit avec MODELLER et qui nous a guidé pour la mutagenèse dirigée;
- (iii) Le modèle décrit par Kosinski et al. et obtenu selon la méthode Frankenstein;
- (iv) Les modèles de chaque domaine C-terminal à l'aide du serveur I-TASSER (Zhang, 2008). Ce serveur est régulièrement classé en tête du concours CASP de prédiction de structures;
- (v) Des modèles supplémentaires à partir de ces précédents modèles en poly-Ala ou en poly-Ser;
- (vi) Des modèles obtenus avec des modes normaux : La structure de ecMutL a montré que chacun des monomères du CTD est constitué de deux sous-domaines (dimérisation et régulateur) susceptibles d'adopter des orientations différentes l'une par rapport à l'autre. Pour tenir compte d'éventuels mouvements de ces sous-domaines, nous avons calculé pour chaque modèle les modes normaux des dix plus basses fréquences à l'aide du serveur *elNémo* (Suhre and Sanejouand, 2004). Les perturbations minimales et maximales sont fixées à -100 et +100 (échelle arbitraire) et calculées par pas de 20 en perturbations. Cela permet ainsi de fournir 110 modèles perturbés pour chaque modèle par homologie.

Ces modèles ont été testés en remplacement moléculaire soit sous la forme de dimères (ecMutL) ou d'hétérodimères (MutL α) des régions CTD. Nous avons également essayé de positionner les modèles des monomères des régions CTD d'ecMutL, Mlh1 ou Pms1. Ces essais (dimères et monomères présentent chacun des avantages différents

- Dimères: Ces modèles représentent une fraction significative de l'unité asymétrique. Si l'interface de dimérisation chez *E. coli* et *S. cerevisiae* est proche, ils constituent un bon modèle.
- Monomère: L'utilisation de la structure cristallographique des monomères de ecMutL ou des différents modèles de scMlh1 et scPms1 permet d'essayer de positionner les monomères l'un après l'autre: Cela présente l'avantage d'autoriser

d'éventuels mouvements à l'interface des monomères. Ces modèles représentent encore une fraction significative de l'unité asymétrique. Avec l'utilisation des monomères, nous pouvons évaluer la qualité des solutions de remplacement moléculaire en fonction des dimères qu'elles permettent ou non de reconstituer.

Le calcul du volume de Matthews et du pourcentage de solvant dans le cristal pour un, deux ou trois complexes montre que l'unité asymétrique peut contenir un complexe avec un pourcentage de solvant élevé (67%) soit deux complexes avec un pourcentage de solvant faible (34%) (**Tableau 3 - 9**).

Tableau 3 - 9 : Evaluation du nombre de molécules dans l'unité asymétrique par le calcul du volume de Matthews ; P(2.60Å) et P(total) représentent les probabilités calculées à partir, respectivement, de structures de résolution égale à 2.60Å et à partir de l'ensemble des structures à partir de 11000 structures cristallographiques non redondantes (Kantardjieff and Rupp, 2003)

Nombre de molécules/ unité asymétrique	Volume de Matthews (Å ³ /Da)	Pourcentage de solvant	P(2.60Å)	P(total)
1	3.75	67	0.30	0.32
2	1.87	34	0.68	0.66
3	1.25	1.6	0.02	0.01

Les essais de remplacement ont été réalisés en parallèle avec les programmes *MolRep* (Vagin and Teplyakov, 2009) et *Phaser* (McCoy et al., 2007). Ces essais ont été réalisés en coupant les données de diffraction à différentes résolutions. Notre jeu de données natives étant à 2.6Å, les essais de remplacement moléculaire ont été réalisés à 2.6Å, 3Å, 4Å, 5Å, 6Å et 7Å. L'avantage des plus hautes résolutions est qu'elles incluent un plus grand nombre d'observations, mais les scores calculés sont plus sensibles aux différences entre le modèle utilisé et la structure présente dans le cristal.

Tous ces tests, comprenant les différents modèles, les différentes méthodes et les différentes résolutions, ont été implémentés de façon automatique au sein de plusieurs scripts³¹. Chaque procédure comprend, en général, l'étape de remplacement moléculaire pour deux dimères ou quatre monomères. Dans le cas du remplacement moléculaire avec *MolRep*, chaque placement d'une entité (un monomère ou un dimère) est suivi d'un affinement en corps rigide effectué par le programme *Buster*, étape qui n'est pas nécessaire avec *Phaser* qui réalise automatiquement une étape d'affinement après avoir placé chaque entité. A la fin du remplacement moléculaire, les solutions sont affinées à l'aide de *Buster* (**Figure 3 - 13(B)**).

Le résultat de cette procédure montre qu'aucune solution ne ressort clairement par rapport aux autres, que l'on considère les valeurs de contraste de *MolRep*, de gain de vraisemblance de *Phaser* ou de R/R_{free} de *Buster*. Concernant ce dernier paramètre,

³¹ Les scripts de remplacement moléculaire pour deux monomères sont présentés en ANNEXE 12

il était compris entre 37 et 40% pour le facteur R et 42 et 46% pour le facteur R_{free}. Cependant, les solutions obtenues présentent certains points communs:

- Nombre de dimères ou monomères placés : Lorsque l'on demande aux programmes de remplacement moléculaire de placer deux dimères ou quatre monomères, ceux-ci n'arrivent à placer qu'un seul dimère ou deux monomères. Les modèles additionnels positionnés présentent de nombreux recouvrements avec les premières molécules positionnées.
- Empilement cristallin : Le positionnement d'un seul dimère correspond comme évoqué ci-dessus à un pourcentage de solvant élevé dans le cristal (70%). Un grand nombre de solutions obtenues avec *MolRep* ou avec *Phaser* présente un empilement cristallin en «mille-feuille» constitué d'alternance entre des tranches de protéines et des tranches de solvant (**Figure 3 - 13(C)**).
- Structure cristallographique versus modèle : Les solutions les plus contrastées obtenues avec *MolRep* ou *Phaser* correspondent aux essais réalisés avec comme modèle de départ la structure cristallographique du CTD d'ecMutL. Malgré différentes approches, la modélisation moléculaire n'a pas permis dans notre étude d'améliorer les résultats de remplacement moléculaire.

A la fin de ces essais de remplacement moléculaire nous avons choisi d'affiner certaines des solutions et d'analyser les cartes de densité électronique. Une des solutions ainsi étudiée correspond à deux monomères du CTD d'ecMutL, placés successivement par *Molrep*. Les deux monomères positionnés reconstituent un dimère avec les deux feuillets des domaines de dimérisation en interaction. Cette solution présente des facteurs initiaux R et R_{free} de 34.0% et 43.8% respectivement qui descendent à 30.8 et 40.1% après quelques cycles d'affinement/reconstruction. Les cartes de densité électroniques obtenues avec ce modèle ou en omettant certaines régions de ce modèle ne sont pas convaincantes. A cette étape du projet, nous avons choisi de réorienter les efforts vers la recherche de dérivés lourds.

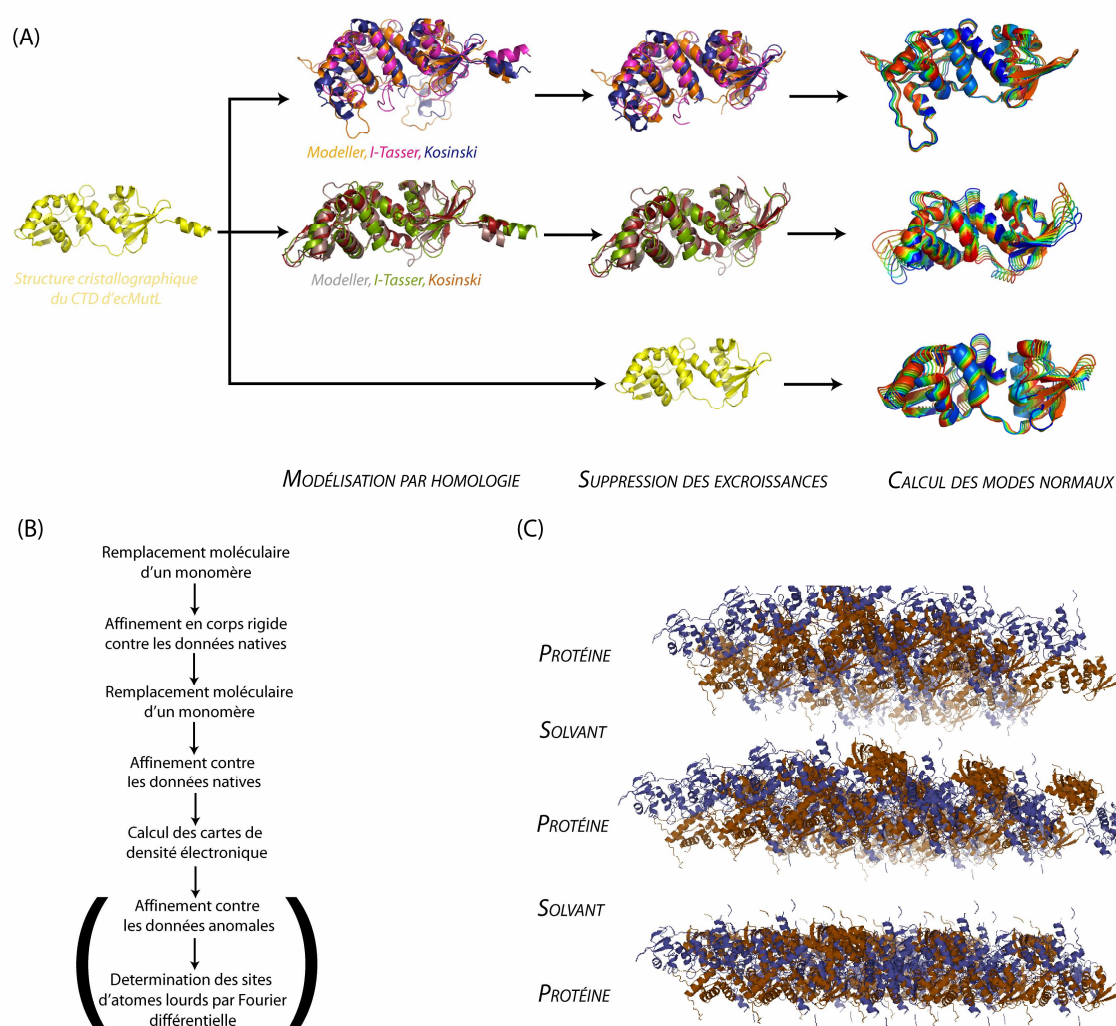


Figure 3 - 13 : Stratégie et résultat du remplacement moléculaire (A) Stratégie de modélisation adéquate avec le remplacement moléculaire (B) Etapes du remplacement moléculaire de deux monomères avec le programme MolRep (C) Empilement cristallin de la solution de remplacement moléculaire obtenu avec MolRep à partir de deux monomères du CTD d'ecMutL

Essais de phasage par la méthode des dérivés lourds

Après les essais infructueux de remplacement moléculaire, nous avons essayé d'introduire des atomes lourds soit par trempage des cristaux du complexe Mlh1*Pms1 dans des sels d'atomes lourds soit directement par co-cristallisation du complexe Mlh1*Pms1 en présence de sels d'atomes lourds.

1- *Ligands potentiels de l'endonucléase Pms1: sels de Mn^{2+} et Zn^{2+} .* Nous avons choisi dans un premier temps de centrer nos efforts sur les cations divalents Mg^{2+} , Mn^{2+} et Zn^{2+} . Les données de la littérature sur Pms1 et son activité endonucléase suggéraient que certains de ces cations pouvaient être des co-facteurs de Pms1. De plus, lors de la cristallisation du complexe Mlh1*Pms1, nous avons observé que l'ajout de 10mM $MgCl_2$ améliore la croissance des cristaux. Le magnésium ne peut

être exploité pour le phasage car il possède un pic d'absorption à trop faible énergie (1.30keV soit $\lambda=9.5\text{\AA}$ pour la transition K) qu'aucune ligne de lumière n'atteint. Les cations Mn^{2+} et Zn^{2+} ont un pic d'absorption à une énergie suffisamment haute pour être accessible sur les lignes de lumière (6.54keV et 9.66keV respectivement, pour la transition K). Les essais par co-cristallisation avec ces deux cations ont montré que les cristaux en présence d'ions Mn^{2+} sont petits et ont un faible pouvoir de diffraction. Aucun cristal n'apparaît en présence de 1mM ZnCl_2 .

Dans un second temps, nous avons centré nos efforts sur les atomes lourds utilisés pour résoudre un nombre important de structure cristallographique. M. Rould présente un classement des sels d'atomes lourds en fonction du nombre de structures qu'ils ont permis de résoudre (Rould, 1997). La majorité des atomes lourds est classée en deux familles : la classe A, incluant les métaux alcalins, les métaux alcalino-terreux (dont Mg^{2+} , Ca^{2+} , Sr^{2+} et Ba^{2+}), les lanthanides et les actinides, qui forment des interactions électrostatiques avec les acides aminés de type aspartate ou glutamate. La classe B inclut des atomes comme l'or, le platine ou encore le mercure. Ces atomes forment fréquemment des liens covalents avec les groupements amines et les thiols des histidines, cystéines ou méthionines.

2- Atomes lourds dits de la classe A: La^{3+} , Eu^{3+} et Sm^{3+} . L'objectif à travers l'utilisation de ces atomes lourds était d'essayer de substituer les ions Mg^{2+} utilisés dans la cristallisation du complexe Mlh1*Pms1 par des atomes lourds de classe A. Nos premiers essais ont porté sur l'utilisation de sels lanthanides qui, bien qu'ayant un degré d'oxydation de III, sont souvent utilisés pour remplacer les ions Mg^{2+} . Nous avons donc testé les ions La^{3+} , Eu^{3+} et Sm^{3+} en co-cristallisation à 1mM, en trempage à 1mM pendant 2 jours et en trempage à 10mM pendant, 5, 10, 15, 20 et 25min. Dans tous les cas, les trempages longs rendent les cristaux « caoutchouteux », nous pouvons tordre les cristaux sans les casser contrairement aux cristaux natifs. Les trempages courts endommagent le cristal car ils réduisent considérablement leur pouvoir de diffraction. Les cristaux trempés dans SmCl_3 et LaCl_3 diffractent au mieux à 7\AA et ceux trempés dans EuCl_3 à 10\AA contre 3.5\AA en moyenne pour les cristaux natifs du complexe Mlh1*Pms1. Les essais de co-cristallisation réalisés en présence de sels de lanthane et de samarium n'ont pas permis d'obtenir de cristaux. Le seul lanthanide pour lequel nous avons obtenu des co-cristaux est un **sel d'euporium(III)**, le chlorure d'euporium. Ces cristaux diffractent jusqu'à 4.5\AA au seuil d'absorption de l'euporium (7keV, $\lambda=1.77\text{\AA}$). L'analyse des données de diffraction montre que les jeux avec l'euporium présentent des différences d'isomorphisme trop importantes par rapport à la native ($R_{\text{iso}}=38,5\%$). De plus, le signal anomal est extrêmement faible (la différence anormale moyenne $|F_+-F_-|/\sigma$ vaut 1.01 ([Tableau 3 - 10](#))).

Les sels de strontium sont, quant à eux, souvent utilisés comme substituant notamment du calcium(II). Nous avons fait des trempages longs et courts comme précédemment ainsi que des essais de co-cristallisation. Dans tous les cas, les cristaux trempés ne diffractent plus. Les cristaux obtenus en co-cristallisation en présence de 1mM SrCl₂, ne diffractent au mieux qu'à 7Å.

Tableau 3 - 10: Statistiques des données de diffraction des dérivés à l'euporium et au mercure et de la native à basse énergie; les valeurs entre parenthèses correspondent aux statistiques dans la dernière coquille de résolution

	Native à basse énergie	Co-cristal à Eu ³⁺	Trempage au Hg ²⁺
Collection des données			
Groupe d'espace	C2	C2	C2
Paramètres de maille			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	187.8, 66.5, 74.4	185.7, 65.8, 74.1	194.2, 67.5, 74.1
α , β , γ (°)	90.0, 91.0, 90.0	90.0, 91.4, 90.0	90.0, 90.3, 90.0
Longueur d'onde (Å)	1.800	1.771	0.9800
Résolution (Å)	40.1-3.70 (3.90-3.70)	39.8-4.50 (4.74-4.50)	33.7-3.98 (4.19-3.98)
<i>R</i> _{sym} or <i>R</i> _{merge}	10.5 (25.7)	11.1 (18.9)	11.8 (49.1)
<i>I</i> / σI	10.2 (5.1)	6.8 (5.6)	4.9 (2.2)
Complétude (%)	99.5 (99.2)	98.8 (99.6)	97.3 (90.4)
Redondance	5.3 (5.3)	3.0 (3.0)	3.1 (3.1)
<i>R</i> _{iso}		38.5 (47.1)	41.9 (54.5)
SigAno	0.843	1.010	0.908
Corr Anomal	8%	35%	26%

Le dernier atome lourd de classe A qui a été testé, est l'uranium(VI) sous forme d'ion uranyle UO₂. Des trempages courts ont été réalisés en présence de 2mM UO₂(Ac)₂. L'ion uranyle est un composé jaune. Nous observons que suite au trempage, les cristaux se colorent légèrement. Parmi les cristaux testés, deux d'entre eux diffractent jusqu'à 5Å.

3- Atomes lourds dits de la classe B: Pt, Hg, Au.

3a. Les sels de platine: Dans un premier temps, nous avons testé trois sels de platine: K₂PtCl₄, K₂PtCl₆ et K₂Pt(CN)₄. Nous n'avons pas réussi à obtenir de co-cristaux. L'ajout de 1mM de ces sels entraîne la précipitation du complexe Mlh1*Pms1 rendant difficile la co-cristallisation. Nous avons effectué deux types de trempages avec les sels de platine: quelques minutes à 10mM ou quelques jours à 1mM. Dans le premier cas, les cristaux ne diffractent plus mais restent fragiles et incolores. Suite aux trempages longs, les cristaux ne diffractent plus également, deviennent élastiques et prennent une teinte orange, couleur du sel de platine utilisé le K₂PtCl₄ (Figure 3 - 14). Cette observation semble indiquer que le platine réagit avec le complexe dans le cristal cependant cette réaction s'accompagne d'une perte trop importante du pouvoir de diffraction des cristaux trempés.

3b. Les sels d'or: Nous avons 2 sels d'or (NaAuCl_4 et $\text{KAu}(\text{CN})_2$) avec une approche similaire à celle utilisée pour les sels de platine. Les sels d'or sont complémentaires du platine car ils sont décrits comme moins réactifs. Dans les mêmes conditions de trempage que celles utilisées pour les sels de platine, les cristaux trempés dans les sels d'or ne diffractent plus. En conclusion, les sels de platine et d'or semblent être des pistes intéressantes de part leur réactivité avec les cristaux de Mlh1*Pms1 mais leur utilisation demande un travail de mise au point important pour maintenir la diffraction des cristaux de Mlh1*Pms1 après trempage.



Figure 3 - 14 : Cristal du domaine C-terminal de Mlh1*Pms1 après un mois de trempage à $1\text{mM K}_2\text{PtCl}_4$

3c. Les sels de mercure: Nous avons testé quatre sels de mercure : HgCl_2 , $\text{Hg}(\text{CN})_2$, $\text{Hg}(\text{Ac})_2$ et CH_3HgCl . Comme pour les sels de platine, la co-cristallisation est limitée par la précipitation du complexe Mlh1*Pms1 en présence de 1mM de sels. Des trempages courts et longs sont menés en parallèle comme pour les sels de platine. Les trempages de 48h à 1mM transforment également les aiguilles en baguettes élastiques. En revanche, c'est parmi les nombreux essais de trempage dans 10mM de sels de mercure que nous avons obtenu les résultats les plus intéressants sur les cristaux de Mlh1*Pms1 par l'approche des dérivés lourds. Des cristaux trempés dans $10\text{mM Hg}(\text{Ac})_2$ pendant 10min diffractent jusqu'à 4\AA . Un jeu de données complet a été collecté au seuil d'absorption du mercure (12.63keV , $\lambda=1.02\text{\AA}$) (**Tableau 3 - 10**). Ces cristaux présentent un signal anomal jusqu'à 6\AA . Les essais de trempage avec les sels de mercure $\text{Hg}(\text{CN})_2$ et CH_3HgCl ont donné des cristaux de pouvoir diffractant plus faible.

Exploitation du jeu de données des cristaux trempés avec $10\text{mM Hg}(\text{Ac})_2$ La carte anormale de Patterson calculée avec le jeu de donnée $\text{Hg}(\text{Ac})_2$ présente des pics potentiellement intéressants. Ces pics sont cependant d'une intensité relativement faible (en dessous de 5σ) par rapport à ce que l'on attend pour un atome de mercure. A l'énergie choisie pour la collecte des données de diffraction, la composante imaginaire f'' de la diffusion anormale du mercure est de $10e$ ce qui correspond à un

signal anomal fort (**Figure 3 - 15(A)&(B)**). Nous avons alors décidé d'utiliser une procédure automatique de recherche des sites d'atomes lourds avec ShelxD à plusieurs résolutions (5, 5.5, 6, 6.5 et 7Å) et en variant le nombre de sites recherchés (4, 6, 8, 10, 12 ou 16 sites). Les sites identifiés par cette procédure sont comparés grâce au programme *Phenix.emma*. On observe alors qu'un certain nombre de solutions sont identiques pour un faible nombre de sites (4, 6 ou 8 sites) et à une résolution de 6 Å. Nous avons alors poursuivi le phasage avec le programme Sharp mais les figures de mérite en sortie de phasage sont très faibles (vers 0.1) pour les deux mains et après aplatissement de solvant avec le programme Parrot, nous obtenons des pouvoirs de phasage de 0.23 pour la main d'origine et 0.30 pour la main inversée. Les cartes résultantes sont alors très bruitées et elles ne semblent pas être un point de départ raisonnable pour une reconstruction. Bien que la délimitation entre les zones de solvant et les zones de protéines soit assez claire, aucun élément de structure secondaire ne ressort de ces cartes.

Enfin, nous avons essayé de combiner ces données anomaes avec les solutions proposées par remplacement moléculaire car, bien qu'insuffisantes pour résoudre expérimentalement les phases, ces données peuvent servir pour valider les solutions de remplacement moléculaire. Nous avons réutilisé les scripts précédents de remplacement moléculaire mais à la suite de l'affinement dans les données natives, la solution de remplacement moléculaire est à nouveau affinée dans les données Hg(Ac)₂. Cette solution est alors utilisée dans le programme *Phaser* pour déterminer, par transformée de Fourier différentielle, la présence de pics de densité pouvant correspondre aux atomes de mercure (routine fournie par P. Legrand). Cependant, aucune des solutions obtenues par remplacement moléculaire ne permet de localiser des sites de forte densité électronique (au-delà de 5 σ), ce qui confirme qu'aucune de ces solutions n'est convaincante.

Analyse a posteriori des essais de dérivés lourds à la lueur de la structure résolue avec le complexe marqué aux sélénométhionines: Après avoir résolu la structure du domaine C-terminal de MutL α (le phasage est détaillé au §3.2.1.*iv* et l'analyse au §3.2.2), nous avons ré-analysé le jeu de données obtenu avec le Hg(Ac)₂. Nous avons résolu par remplacement moléculaire la structure du CTD de Mlh1*Pms1 dérivé avec le Hg(Ac)₂. L'analyse des cartes de différence anomaes entre ce dérivé et la structure native du CTD de Mlh1*Pms1 révèle quatre pics supérieurs à 5 σ : deux pics correspondent aux deux sites métalliques du site endonucléase de Pms1; les deux autres pics sont à proximité de groupements sulfures. Lorsque l'on affine la structure du dérivé avec un atome de mercure à ces 2 dernières positions, les pics positifs de la carte différentielle F_o-F_c disparaissent. Cohérent avec l'affinité entre les ions mercure(II) et les groupements thiols et amines, un des mercures se lie à la cystéine C738 à la surface de Mlh1 à proximité des arginines R741 et R742. Le dernier

mercure est lié à la cystéine C685 à proximité des acides aminés M623 et M626. (Figure 3 - 15(C)).

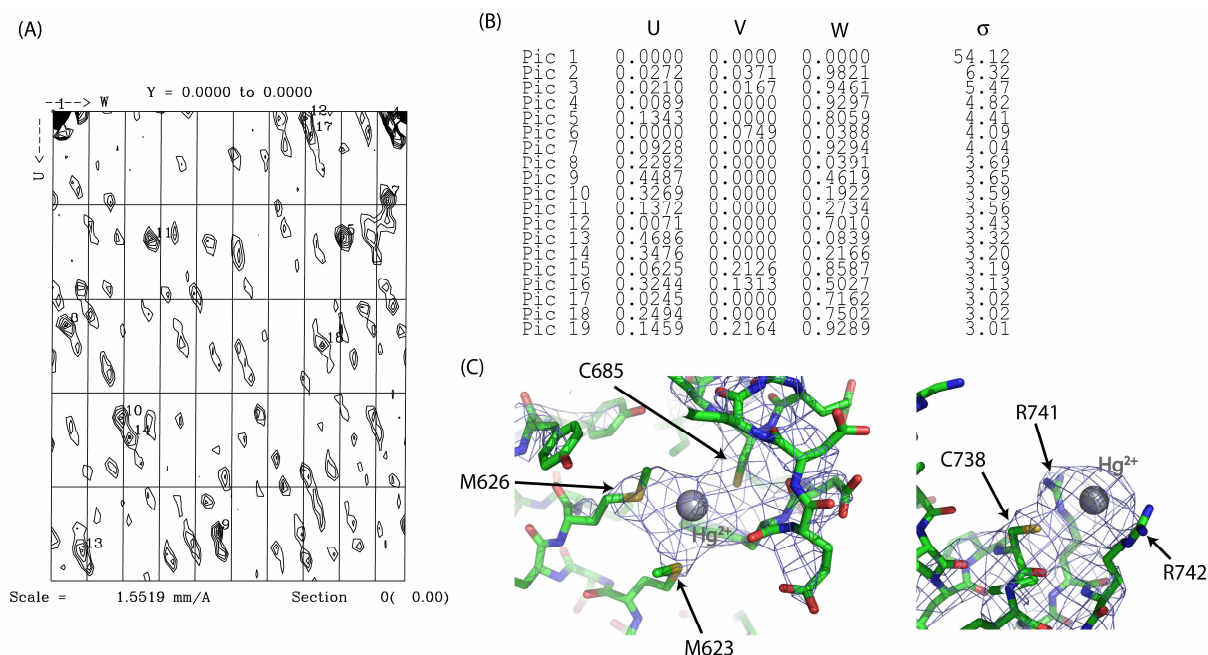


Figure 3 - 15 : Analyse du dérivé Hg(Ac)₂ (A) Carte de Patterson utilisant la différence anormale suivant les axes U(0-0.5) et W(0-1) ; le paramètre V ne peut être déterminé dans le groupe d'espace C2 (B) Pics principaux de la carte de Patterson anormale (pics supérieurs à 3σ) (C) Cartes de densité électronique 2F_o - F_c à un contour de 1.2σ calculé à partir du remplacement moléculaire du dérivé Hg(Ac)₂ avec la structure du CTD de Mlh1*Pms1 résolue ultérieurement. Les vues représentent deux sites probables de fixation des ions Hg²⁺ à la surface du CTD de Mlh1

Essais de phasage avec les atomes de soufre : En parallèle de ces deux premières approches classiques de phasage par remplacement moléculaire et dérivées lourds, nous avons également souhaité tester la méthode de phasage qui utilise le très faible signal anormal du soufre des méthionines et des cystéines des protéines du cristal. Pour cela, nous avons enregistré un jeu de données à 3.7Å à basse énergie (7keV, $\lambda=1.77\text{\AA}$) sur le même cristal avec lequel nous avons enregistré les données natives. Cependant, à cette énergie, la composante imaginaire f'' de diffusion anormale du soufre est faible (0.8e). Cette approche nécessite une grande redondance des données pour mesurer précisément de petites différences anormales. Or le jeu de données enregistré à basse énergie ne présente qu'une redondance de 5.3 avant que les dommages du cristal ne soient trop importants, là où une redondance d'au moins 20 aurait été requise (Tableau 3 - 10). Les jeux de données enregistrés n'ont pas permis au final de déterminer la position des atomes de soufre et de calculer des premières phases.

En conclusion, un des obstacles principaux rencontré lors des essais de dérivés lourds a été le très faible pouvoir diffractant et la sensibilité aux rayons X de la majorité des cristaux du CTD de Mlh1*Pms1 mis en présence de sels d'atomes lourds.

Les résultats les plus encourageants ont été obtenus avec les sels de mercure et d'uranyle. Cependant, à l'issue de plusieurs mois passés sur la recherche de dérivés lourds, aucune phase de qualité suffisante pour commencer à construire un modèle du CTD de Mlh1*Pms1 n'a été obtenue. Nous avons alors choisi de réorienter nos efforts vers le marquage du complexe avec des sélénométhionines.

iv. Détermination des phases par marquage du CTD de Mlh1*Pms1 avec des sélénométhionines

*Expression et purification du domaine C-terminal de Mlh1*Pms1 marqué avec des sélénométhionines*

Historiquement, dès l'obtention d'un jeu de données natives à une résolution meilleure que 3.0Å, nous avons réalisé au laboratoire, les premiers essais de purification du complexe Mlh1*Pms1 marqué avec des sélénométhionines. Aucun de ces premiers essais n'a permis d'obtenir du complexe purifié à homogénéité et en quantité suffisante. Nous avons alors choisi de recentrer sur une durée de six mois nos efforts sur le remplacement moléculaire et les dérivés lourds. Devant les difficultés rencontrées avec ces deux approches, comme évoqué dans les paragraphes précédents, nous avons décidé de reprendre les essais de marquage avec des sélénométhionines.

Le protocole de marquage du complexe avec des sélénométhionines que nous avons choisi est un protocole qui nous a été communiqué par K. Blondeau (IBBMC). Ce protocole, détaillé dans le chapitre "Matériels et Méthodes", utilise une expression en milieu minimum avec ajout de sélénométhionines et sans recours à des souches auxotrophes. Comme pour le complexe natif, le complexe est co-exprimé par transformation par le plasmide MBP-Pms1, de cellules BL21 compétentes, contenant le plasmide GST-Mlh1. Les précultures réalisées en milieu LB sont introduites dans du milieu minimum. Dans les premiers essais, l'expression du complexe marqué avec des sélénométhionines était plus faible que celle observée avec le complexe natif. Nous nous sommes aperçus que ce taux d'expression n'était pas lié au milieu minimum mais à la fraîcheur des cellules compétentes. L'utilisation de bactéries compétentes fraîches (stockées moins de deux mois à -80°C) a permis d'améliorer l'expression.

Le premier protocole de lyse utilisé était celui utilisé au départ pour le complexe natif. Lors des premières productions, il s'est avéré qu'une forte proportion du complexe marqué était retrouvée dans la fraction insoluble. Nous avons choisi d'essayer des tampons de lyse différents. Nous avons choisi notamment d'ajouter 0,5% (v/v) de Triton, 5% (v/v) de glycérol et 500mM de NaCl (communication F. Ochsenbein). Cette modification a permis de récupérer une fraction plus grande de complexe dans la fraction soluble.

Le protocole de purification du complexe marqué est identique à celui du complexe natif, à savoir: (i) une étape de capture sur résine amylose, (ii) le clivage des étiquettes par la protéase TEV, (iii) une étape d'affinité sur colonne de nickel et

enfin (iv) une étape de polissage sur échangeuse anionique. Ce protocole permet d'obtenir 10mg environ du complexe GST-Mlh1/MBP-Pms1 dans l'éluat de la résine amylose par litre de culture d'*E. coli*. A l'issue de l'étape de polissage, nous obtenons environ 1mg de complexe Mlh1*Pms1 par litre d'*E. coli* avec un bon degré de pureté (**Figure 3 - 16(A)**).

*Cristallisation du domaine C-terminal de Mlh1*Pms1 marqué avec des sélénométhionines*

Les essais de cristallogénèse ont été réalisés autour des conditions identifiées pour le complexe natif. De fines aiguilles apparaissent en présence de MPEG 5k dans une gamme de 5 à 10%(m/v) et 0.1M HEPES pH7 ou MES pH6.5. Dans le cas du complexe marqué à la sélénométhionine (SeMet), ces cristaux n'apparaissent qu'à 4°C alors que les cristaux natifs peuvent être obtenus à 4°C ou 20°C. Une autre différence notable peut être observée entre les cristaux natifs et les cristaux SeMet, ces derniers apparaissant au milieu de précipités de protéines. L'ajout d'ion Mg^{2+} ne permet pas d'améliorer la taille des cristaux comme cela était observé pour le complexe natif dans une gamme de concentration de 1 à 100mM. Les aiguilles de cristaux SeMet sont plus petites que celles du complexe natif (largeur et épaisseur inférieures à 5 μ m). Malgré leur petite taille (dimension 100 \times 5 \times 2 μ m³), les premiers cristaux SeMet testés ont un pouvoir de diffraction prometteur puisqu'un cristal de cette taille a permis d'obtenir un premier jeu complet à 5Å (**Figure 3 - 16(B)**).

Afin d'obtenir des cristaux de taille supérieure, nous avons choisi de réaliser des expériences d'ensemencement. Dans le cas de la cristallisation du complexe natif, l'ensemencement avait permis de travailler dans des conditions où les cristaux n'apparaissaient pas spontanément, ni de façon trop importante. Dans le cas de la cristallisation du complexe SeMet, l'ensemencement a permis l'inverse à savoir l'augmentation du nombre de gouttes de cristallisation avec des cristaux de SeMet dans la mesure où ces cristaux apparaissent rarement spontanément. Les essais d'ensemencement sont réalisés en apportant des microcristaux broyés ou des grains d'hydroxyapatite dans des gouttes de complexe SeMet. Le contrôle de la croissance a permis de gagner en taille sur la largeur et l'épaisseur des cristaux; un cristal a notamment été obtenu avec une taille supérieure (200 \times 15 \times 6 μ m³) (**Figure 3 - 16(C)**).

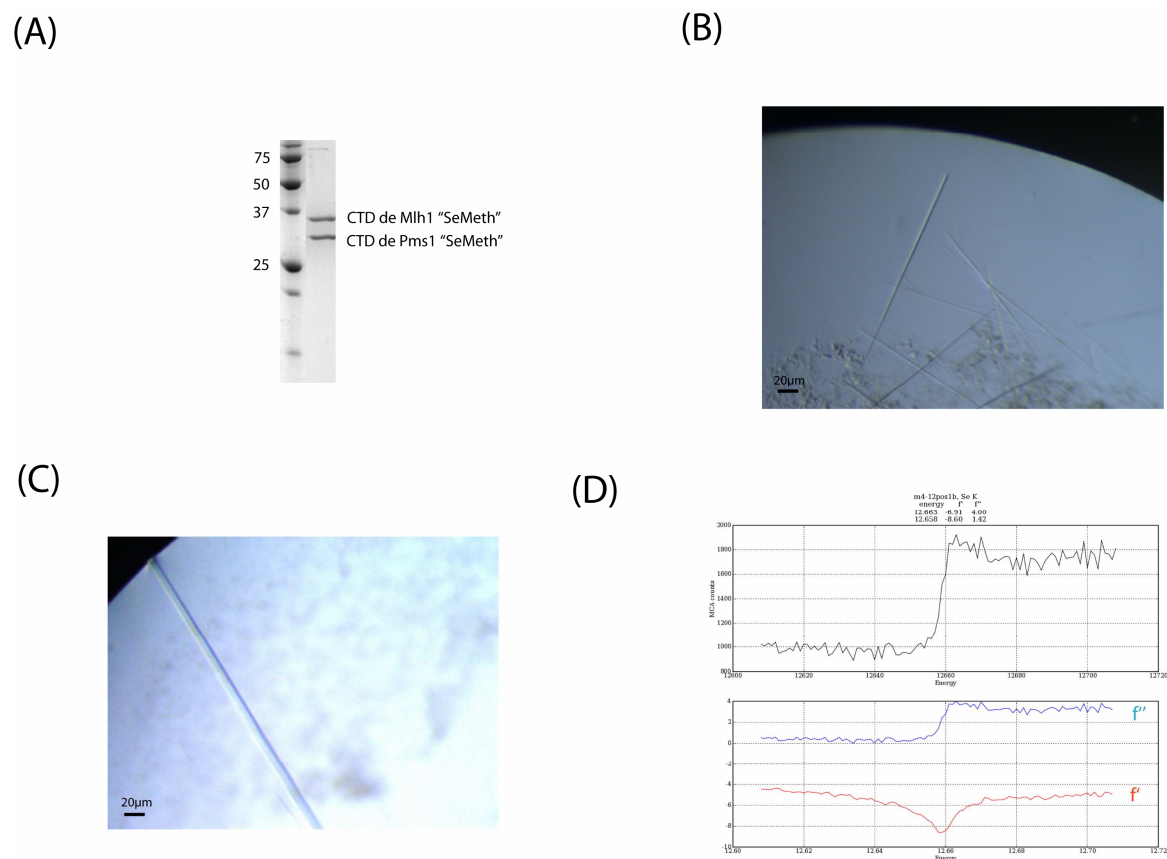


Figure 3 - 16 : Production de cristaux du domaine C-terminal de MutL α substitué au sélénium (A) Analyse électrophorétique sur gel d'acrylamide en conditions dénaturantes en sortie de colonne échangeuse d'anions du domaine C-terminal de Mlh1 substitué au sélénium (B) Cristaux obtenus par nucléation spontanée à 4°C; la solution du réservoir contient 8%(w/v)MPEG 5K ; 0.1M HEPES pH7 ; la solution du complexe protéique a une concentration de 5 mg/ml (C) Cristal obtenu par ensemencement d'hydroxyapatite à 4°C; la solution du réservoir contient 9%(w/v)MPEG 5K ; 0.1M MES pH6.5 ; 10mM MgSO₄ ; la solution du complexe protéique a une concentration de 5 mg/ml (D) Spectre de fluorescence du cristal précédent autour de l'énergie d'absorption de la transition K du sélénium

Détermination de la position des sur-structures et des phases expérimentales

Une cinquantaine de cristaux SeMet ont été analysés sur la ligne Proxima1. La plupart de ces cristaux diffracte à moins de 5 Å. Seul le cristal de taille supérieure évoqué dans le paragraphe précédent a permis d'obtenir un jeu complet à 3.1Å de résolution au pic d'absorption du sélénium (12.65keV, $\lambda=0.98\text{\AA}$). Pour s'assurer que les sélénométhionines sont bien incorporées dans le complexe Mlh1*Pms1, un spectre de fluorescence autour du seuil d'absorption de la transition K du sélénium (12.65keV en théorie) a été réalisé. Ce spectre montre bien une absorption à une énergie de 12.66keV (**Figure 3 - 16(D)**).

Nous avons choisi d'optimiser la collecte d'un seul jeu de données au pic d'absorption du sélénium et donc d'essayer de déterminer les phases par la méthode SAD (Single wavelength Anomalous Dispersion). Les nombreuses collectes réalisées sur les cristaux natifs ont en effet montré que les aiguilles du complexe du CTD de Mlh1*Pms1 se dégradent rapidement. Il semblait donc difficile de réaliser trois jeux

de données (au pic, à l'inflexion et au-delà du pic) sur un même cristal pour utiliser la méthode MAD (Multi-wavelength Anomalous Dispersion). Le SAD semblait également une option intéressante de par le fort pourcentage de solvant de nos cristaux. Cette caractéristique est favorable pour déterminer la bonne main des phases, entre celle d'origine et inverse grâce à l'étape d'aplatissement de solvant.

L'analyse des statistiques d'intégration du jeu de donnée montre un rapport I/σ_I égal 3.7 dans la dernière coquille de résolution, de 3.3 à 3.1 Å (**Tableau 3 - 11**). La différence anormale moyenne est supérieure à 1 jusqu'à 4.5 Å, avec une corrélation anormale supérieure à 30%. Ces données suggèrent un signal anormal faible mais potentiellement intéressant.

Tableau 3 - 11 : Statistiques des données de diffraction du cristal substitué au sélénium; les valeurs entre parenthèses correspondent aux statistiques dans la dernière coquille de résolution

	SeMet Pic du Se
Collection des données	
Groupe d'espace	C2
Paramètres de maille	
a, b, c (Å)	190.9, 66.2, 74.5
α, β, γ (°)	90.0, 90.8, 90.0
Longueur d'onde (Å)	0.9791
Résolution (Å)	45.0-3.12 (3.31-3.12)
R_{sym} or R_{merge}	16.1 (61.0)
$I / \sigma I$	12.3 (3.70)
Complétude (%)	98.7 (96.0)
Redondance	9.0 (9.0)
SigAno	1.068
Corr Anomal	29%

Tableau 3 - 12 : Analyse du signal anormal du cristal SeMet ; la corrélation anormale représente la qualité du signal anormal, la différence anormale moyenne SigAno représente la force du signal anormal

Limite de résolution (Å)	Corrélation anormale (%)	SigAno	Nombre de réflexions utilisées
9.14	85	2.909	568
6.55	63	1.818	1050
5.37	38	1.243	1375
4.66	29	1.127	1603
4.17	17	1.000	1828
3.81	17	0.955	2040
3.53	10	0.874	2209
3.31	6	0.794	2378
3.12	5	0.763	2412
Total	29	1.068	15463

Nous avons alors cherché à déterminer la localisation des atomes lourds avec le programme ShelxD. Le complexe CTD de Mlh1*Pms1 contient 11 méthionines

d'après l'analyse de la séquence des deux régions C-terminales; nous avons donc cherché 20 sites de diffuseurs anomaux. Ceci permet d'englober toutes les méthionines et d'éventuels diffuseurs anomaux autres que le sélénium, comme des ions métalliques ou des atomes de soufre de cystéines. Nous avons réalisé 40 000 essais par méthode directe avec déconvolution de Patterson et à différentes résolutions (4.0, 4.3, 4.5, 4.8 5.0, 5.2, 5.5 et 6.0Å). On observe, pour chaque résolution, un petit nombre de solutions ayant une figure de mérite (FOM) supérieure à 40%. Ces valeurs sont élevées pour la méthode SAD où l'on considère qu'une solution probable a une figure de mérite supérieure à 30%. La très grande majorité des autres solutions a une figure de mérite comprise entre 20 et 30%. Nous avons comparé les 20 sites déterminés avec des FOM supérieures à 40% avec le programme SITCOM (Dall'Antonia and Schneider, 2006). On observe, que pour une résolution donnée, 10 ± 2 sites sont communs à toutes les solutions à haute figure de mérite, sites ayant les occupations les plus proches de 1 (**Figure 3 - 17(A)**). Enfin, ces quelques solutions sont également celles qui ont les plus hauts coefficients de corrélation entre les facteurs de structures calculés et ceux observés, que les coefficients de corrélation soient calculés sur toutes les réflexions (CC_{all}) ou seulement sur les réflexions les plus faibles, non utilisées pour déterminer les sites des sur-structures (CC_{weak}). Les graphes du coefficient de corrélation CC_{all} en fonction du nombre d'essais et du CC_{all} en fonction du CC_{weak} , montrent que les meilleures solutions se détachent nettement des autres solutions (**Figure 3 - 17(B)&(C)**).

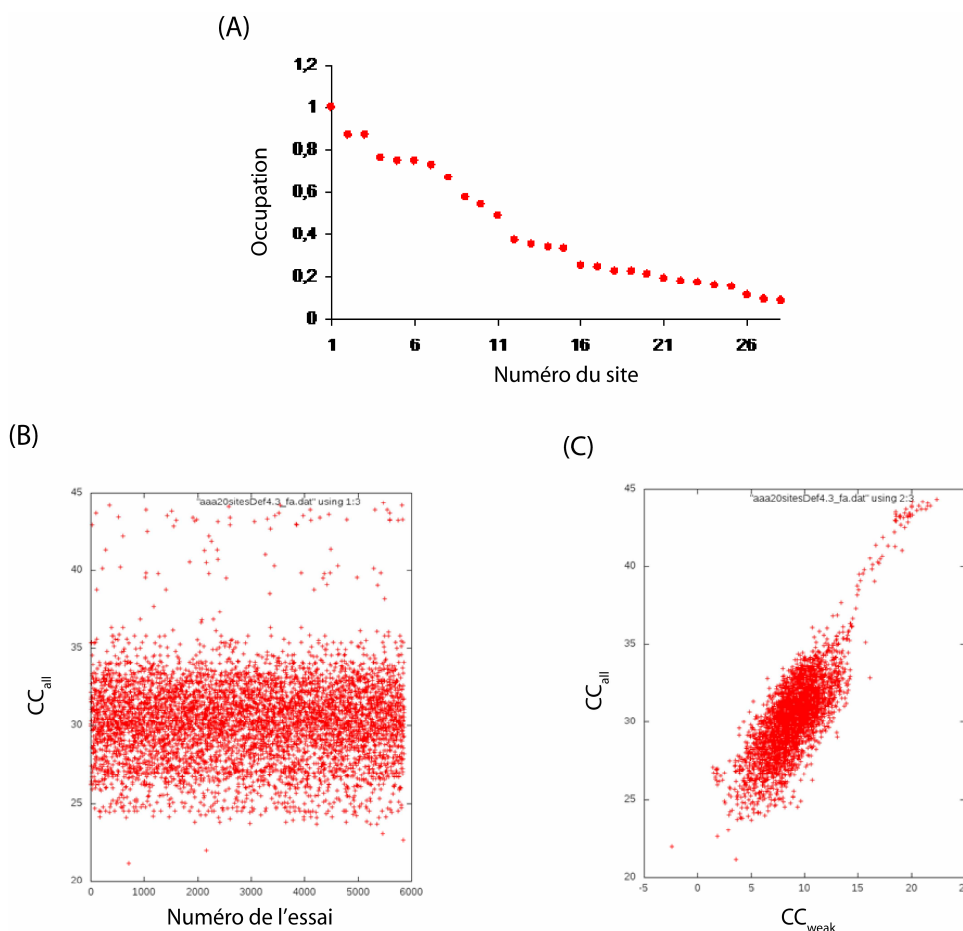


Figure 3 - 17 : Détermination des sites d'atomes lourds par ShelxD (A) Occupation des 28 positions de sur-structures à une résolution de 4.3Å (B) Coefficient de corrélation totale pour les 6 000 premiers essais de 20 sites à une résolution de 4.3Å (C) Coefficient de corrélation total en fonction du coefficient de corrélation des plus faibles réflexions pour déterminer 20 sites à une résolution de 4.3Å

Des 28 sites obtenus dans ShelxD à une résolution de 4.3Å, seuls sont gardés ceux ayant une occupation supérieure à 0.3, c'est-à-dire 15 sites. Ces positions sont affinées et mises à jour par le programme *Phaser* afin de sélectionner les positions qui ne se superposent pas et ont des Z-scores élevés (le Z-score exprime le contraste entre une position et l'estimation du bruit). Le programme *Phaser* calcule ensuite des phases pour les mains d'origine et inversée en utilisant un pourcentage de solvant de 69%. A cette étape, les statistiques de phasage ne sont pas très différentes pour les deux mains. La figure de mérite sur l'ensemble des réflexions est de 30.6% pour la main d'origine et 30.7% pour la main inversée. On remarque également que la figure de mérite est différente selon les coquilles de résolution. Elle est comprise entre 40 et 50% aux basses résolutions et vaut environ 20% entre 4.0 et 3.1Å.

La présence de 70% de solvant dans les cristaux SeMet du complexe Mlh1*Pms1 permet d'améliorer les phases en appliquant plusieurs cycles d'aplatissement de solvant. 20 cycles au total sont réalisés à l'aide du programme Parrot. Les différences entre les deux mains ressortent à la suite de cette procédure. Les phases calculées avec la main d'origine ont un gain de vraisemblance (LLG) de

11054 contre un gain de vraisemblance (LLG) de 6833 pour les phases calculées avec la main inverse. Cette différence se retrouve au niveau du facteur σ_A de probabilité qui permet d'estimer la qualité des phases après aplatissage de solvant (**Tableau 3 - 13**). L'analyse des cartes de densité électronique calculées avec la main d'origine suggère que la structure des cristaux SeMet du complexe Mlh1*Pms1 est résolue. En effet, dans ces cartes, on peut distinguer clairement la frontière entre les régions occupées par le solvant et celles occupées par les protéines. On distingue clairement des éléments de structures secondaires. Les cartes de densité électronique calculées avec la main inverse ne présentent pas ces caractéristiques.

Tableau 3 - 13 : Estimation des probabilités des phases après aplatissage de solvant par Parrot et pour les deux mains

Résolution (Å)	$\sigma_{A, \text{origine}}$	$\sigma_{A, \text{inverse}}$
7.25	0.904	0.760
5.00	0.955	0.835
4.19	0.942	0.889
3.75	0.912	0.853
3.45	0.915	0.886
3.21	0.854	0.838

L'analyse des cartes de densité électronique correspondant à ces premières phases expérimentales permet d'identifier huit hélices (**Figure 3 - 18(A)**). Ces cartes confirment la forte proportion de solvant dans les cristaux SeMet (**Figure 3 - 18(B)**). Nous avons utilisé le programme *Buccaneer* pour effectuer 3 cycles de reconstruction automatique en alternance avec 10 cycles d'affinement avec *Refmac*. Cette procédure permet de placer la chaîne principale de 429 résidus sur 529 résidus du complexe du CTD de Mlh1*Pms1 cristallisé. Le programme *Buccaneer* reconstruit également automatiquement un nombre important de chaînes latérales (319 sur 529). Le programme utilise pour cela les séquences de Mlh1 et de Pms1 qui lui sont fournies.

Afin d'affiner la position, l'occupation et le facteur de température des diffuseurs anomaux, nous avons combiné les phases des sur-structures déterminées précédemment avec celles calculées à partir des premiers modèles issus de la construction automatique. Le programme *Phaser* permet de déterminer 14 positions de sur-structures à partir des solutions initialement proposées par *ShelxD* et de calculer des phases avec une figure de mérite de Molecular Replacement SAD (MRSAD) de 66.9%. Les 14 positions ont des occupations supérieures à 0.6 sauf une qui a une occupation de 0.22. Dans le but de déterminer si cette position correspond bien à un diffuseur anormal, nous avons analysé les cartes de différence anormale de Fourier dans *Coot*, *i.e.* la différence anormale ΔF_{ano} combinée aux phases du modèle déphasé de $\pi/2$ ($\phi_{\text{model}} + \pi/2$). Le site d'occupation 0.22 se trouve à l'emplacement d'une cystéine de Mlh1 et correspondrait donc au signal de l'atome de soufre. Sur les

13 autres sites de diffusion anormale, 11 correspondent à des positions où la construction automatique a placé des méthionines. Les 2 derniers sites, très proches l'un de l'autre, sont situés à proximité d'un des motifs du site endonucléase de Pms1. Ces sites pourraient correspondre aux cations naturels du site endonucléase de Pms1 ou à des cations s'étant substitués aux cations naturels. Comme évoqué en introduction, différentes propositions ont été formulées sur la nature des cations présents au niveau du site endonucléase de Pms1 : Mg^{2+} ou Mn^{2+} d'après la stimulation de l'activité endonucléase (Kadyrov et al., 2007), Zn^{2+} ou Fe^{2+} , d'après les analyses de séquence et de spectrométrie de masse (Kosinski et al., 2008). L'étude de la nature de ces métaux sera détaillée au §3.2.2.*iv*; ces deux positions seront appelées Me pour le moment.

Au final, 14 sites de diffuseurs anomaux ont été identifiés: 11 atomes de sélénium correspondant aux sélénométhionines introduites lors du marquage, 1 soufre d'une cystéine et 2 ions métalliques (**Figure 3 - 18(C)**). Ayant identifié la nature de 12 de ces 14 diffuseurs anomaux, de nouvelles phases expérimentales sont calculées en combinant les phases expérimentales et les phases provenant du modèle reconstruit. Ces nouvelles phases ont une figure de mérite en MRSAD de 67.0%. Si l'on utilise uniquement les sites de diffuseurs anomaux, la figure de mérite en SAD est de 35.8%, donc 5% supérieure à la figure de mérite des phases initiales en sortie de *ShelxD* (FOM=30.6%). L'amélioration de la figure de mérite est particulièrement importante à haute résolution. Les figures de mérite des dernières phases atteignent 30% contre 20% pour les phases déterminées à partir des sites d'atomes lourds de *ShelxD*.

L'étape de construction automatique est effectuée une nouvelle fois avec 3 cycles de *Buccaneer* alternés avec 10 cycles d'affinement avec *Refmac*, cette séquence étant répétée 100 fois. Au cours de cette étape, 462 résidus sont construits sur dix chaînes différentes, une de ces chaînes comportant 261 résidus. A l'issue de cette première phase de reconstruction automatique, 87% de la chaîne principale du CTD du complexe Mlh1*Pms1 est modélisé (462 résidus sur 529 résidus). 81% des chaînes latérales du complexe sont modélisés automatiquement (428 sur 529). L'analyse du modèle reconstruit révèle que le CTD de Mlh1 est celui qui a été le plus reconstruit automatiquement (chaîne principale de 261 résidus contre 201 pour Pms1). De plus, le CTD de Mlh1 est reconstruit sous la forme d'une chaîne continue de 261 acides aminés contre neuf chaînes pour les 168 acides aminés reconstruits de Pms1 (**Figure 3 - 18(D)**).

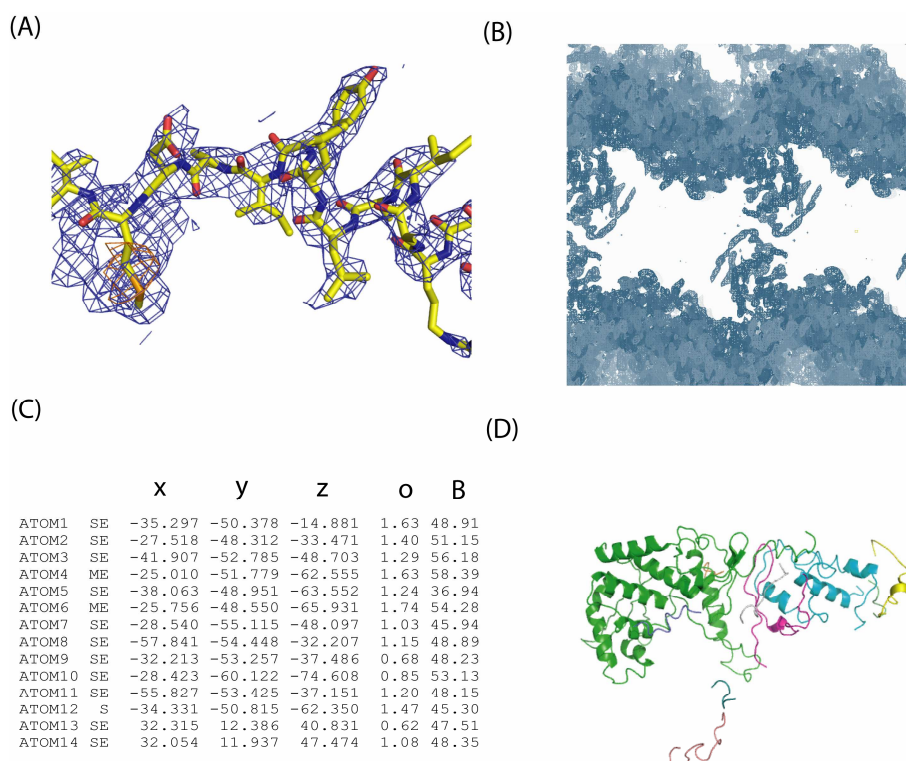


Figure 3 - 18 : Détermination des phases expérimentales et construction automatique (A) Hélice construite automatiquement par *Buccaneer* dans les phases expérimentales (grillage bleu) après aplatissement de solvant ; la différence anormale observée sur une méthionine est représentée en orange (B) Vue générale de la densité électronique (grillage bleu) à partir des phases expérimentales après aplatissement de solvant (C) Sites de diffuseurs anomaux déterminés par MRSAD après 3 cycles de construction automatique (D) Modèle construit dans les phases expérimentales après 100 cycles de construction automatique par *Buccaneer*

*Affinement et validation de la structure du domaine C-terminal de Mlh1*Pms1*

Plusieurs cycles de reconstruction avec le programme *Coot* et d'affinement avec le programme *Buster* sont réalisés dans un premier temps avec le jeu de données du cristal SeMet à 3.1Å. A l'issue de 38 cycles de reconstruction/affinement à cette résolution, la structure présente des facteurs d'accord R et R_{free} de 20.4% et 24.0% respectivement.

A cette étape, la structure est affinée à 2.6Å avec le jeu de données natif en gardant le même échantillon de réflexion pour calculer le R_{free} . Nous avons choisi d'introduire des tenseurs appelés TLS (Translation, Libration and Screw-rotation; macro «*TLSalternate*» de *Buster*) afin de rendre compte des déplacements en corps rigides de différents fragments du complexe Mlh1*Pms1. Les premiers affinements sont réalisés en définissant chaque chaîne comme un segment de TLS. Pour les derniers affinements, nous avons divisé chaque monomère en quatre segments, en tenant compte de la structure des CTD de Mlh1 et Pms1 ainsi que des propositions des programmes qui calculent les segments TLS potentiels (serveur TLS Motion Determination). L'analyse de l'erreur résiduelle des paramètres de déplacement atomique calculés et observés en fonction du nombre de segment de TLS montre qu'au-delà de 4-5 segments, l'erreur résiduelle décroît plus lentement, et ce, pour les

deux monomères (**Figure 3 - 19(A)&(B)**). Les quatre segments proposés par TLSMD sont par ailleurs en accord avec les divisions en sous-domaines décrites sur le domaine C-terminal de MutL ou observées dans cette étude sur les domaines C-terminaux de Mlh1 et Pms1. Deux segments correspondent au domaine de dimérisation de Mlh1 et Pms1 et un segment à leur domaine régulateur. Le segment supplémentaire du monomère de Mlh1 correspond à une insertion de 16 acides aminés (appelée insertion 2). Cette insertion est constituée d'une hélice et d'une boucle et présente des facteurs de température plus importants que le reste de la protéine Mlh1. Quant au domaine supplémentaire du monomère Pms1, il correspond aux derniers résidus de l'extrémité C-terminale qui adopte une structure étendue. L'utilisation de ces segments de TLS s'accompagne d'une diminution de la valeur du R_{free} et d'une amélioration des cartes de densité électronique $2mF_o - F_c$. En fin d'affinement, nous obtenons avec le programme Buster des facteurs d'accord R et R_{free} à 2.6Å de 17.3% et 21.0% respectivement, c'est-à-dire bon pour une structure à 2.6Å (**Tableau 3 - 14**).

Tableau 3 - 14 : Statistiques d'affinement de la structure cristallographique du domaine C-terminal du complexe Mlh1*Pms1 ; les valeurs sont issues du dernier cycle d'affinement avec *Buster*

	<i>Buster</i>
Affinement	
Résolution (Å)	2.60
No. reflections	27 838
R / R_{free}	17.3 / 21.0
No. atomes	
Protéine	3791
Ligand/ion	76
Eau	148
Facteurs B	
Protéine	75.3
Ligand/ion	72.7
Eau	56.2
R.m.s deviations	
Longueurs de liaisons (Å)	0.010
Angles des liaisons (°)	1.170

Le modèle final de la région C-terminale du complexe Mlh1*Pms1 contient 483 acides aminés soit 91% des acides aminés présents dans les constructions utilisées en cristallisation, à savoir, Mlh1 (485-769) et Pms1 (635-873). Les résidus de Mlh1 et Pms1 non modélisés sont ceux situés à l'extrémité N-terminale (acides aminés 485 à 509 pour Mlh1 et 635 à 650 pour Pms1). Parmi les 483 résidus présents dans la structure finale, nous avons pu modéliser dans les cartes de densité électronique les chaînes latérales de 90% d'entre eux (**Figure 3 - 19(C)**). Les résidus pour lesquels seule la chaîne principale a été modélisée sont sur Mlh1 pour sept d'entre eux et sur Pms1 pour les 44 autres. 35 sur les 44 résidus de Pms1 font partie d'un feuillet en brins β de la région régulatrice de Pms1.

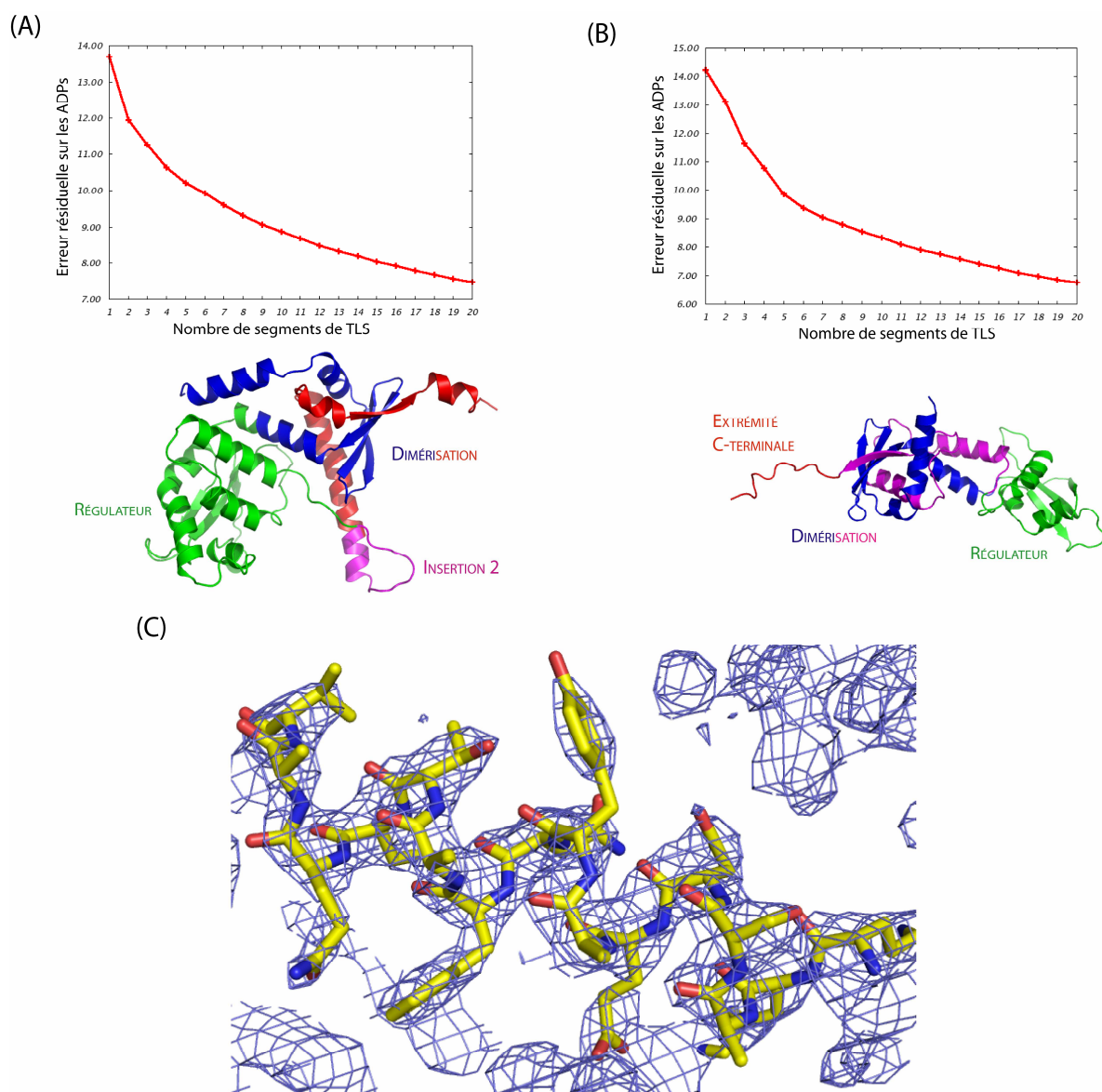


Figure 3 - 19 : Affinement de la structure du domaine C-terminal du complexe Mlh1*Pms1 (A) Erreur résiduelle des ADPs (Atomic Displacement Parameters) et définition des segments de TLS de la région C-terminale de Mlh1 ; les segments de Mlh1 sont respectivement colorés en bleu, vert, magenta et rouge dans l'ordre de la séquence (B) Erreur résiduelle des ADPs et définition des segments de TLS de la région C-terminale de Pms1 ; les segments de Mlh1 sont respectivement colorés en bleu, vert, magenta et rouge dans l'ordre de la séquence (C) Densité électronique $2mF_o - F_c$ de l'hélice αA de Pms1 avec un contour de 2σ

Analyse de la qualité du modèle final : Le modèle final de la région C-terminale du complexe Mlh1*Pms1 a été analysé à l'aide du programme *Model_vs_data*. Les facteurs R et R_{free} estimés par ce programme sont de 18.9% et 23.2% respectivement. L'analyse de la géométrie avec le programme *MolProbity* montre que 96 et 99% des résidus sont respectivement dans les régions les plus favorables et favorables du diagramme de Ramachandran. La structure du CTD de Mlh1*Pms1 se trouve selon les critères de programme parmi les meilleures structures sur un échantillon de 227 structures analysées et résolues entre 2.35Å et 2.95Å (Tableau 3 - 15).

Tableau 3 - 15 : Statistiques de la géométrie de la structure cristallographique du domaine C-terminal du complexe Mlh1*Pms1 ; le score d'inter-pénétration correspond au nombre d'inter-pénétrations stériques >0.4Å pour 1000 résidus ; la librairie des paramètres géométriques est issue de celle d'Engh&Huber, 2001 (Engh and Huber, 2001)

Statistiques de MolProbity		
Contacts atomiques		
Score d'inter-pénétration (Clashscore)	10.97	97 ^e centile (N=227, 2.60Å±0.25Å)
Géométrie du complexe		
Conformations « non-rotamériques »	4.74%	
Hors région autorisée du Ramachandran	0%	
Région favorable du Ramachandran	100%	
Région la plus favorable du Ramachandran	95.83%	
Déviations Cβ >0.25Å	0	
Score MolProbity	2.35	89 ^e centile (N=6237, 2.60Å±0.25Å)
Résidus hors distance autorisée de liaisons	0.00%	
Résidus hors angle autorisé	0.00%	

3.2.3 ANALYSE DE LA STRUCTURE CRISTALLOGRAPHIQUE DU DOMAINE C-TERMINAL DE MUTLα DE *S. CEREVISIAE*

Cette partie est présentée sous une forme "Résultats-Discussion" pour faciliter la lecture et présenter la discussion près des figures auxquelles elle fait référence.

i. Etude globale de la structure cristallographique de domaine C-terminal de MutLα

Extrémités N-terminales des CTD de Mlh1 et Pms1. Les 25 premiers acides aminés de Mlh1 (485-509) et les 16 premiers acides aminés de Pms1 (635-650) n'ont pu être modélisés comme évoqué dans le paragraphe précédent. L'analyse de l'empilement cristallin montre que les premiers acides aminés de Mlh1 et Pms1 modélisés, respectivement N510^M et E651^P, sont situés à proximité de larges cavités de solvant (les indices M et P en exposants après un acide aminé indiquent, par la

suite, des acides aminés de Mlh1 et Pms1, respectivement). Ces positions sont compatibles avec la présence de respectivement 25 et 16 acides aminés non structurés dans le cristal. Cette observation suggère que les acides aminés (485-509) de Mlh1 et (635-650) de Pms1 font partie des longues régions charnières qui relient les domaines N-terminaux et C-terminaux de Mlh1 et Pms1. Notre structure et celles des domaines N-terminaux de Mlh1 et Pms1 permettent de proposer comme limites de ces régions charnières les acides aminés 335 à 507 pour Mlh1 de levure et 395 à 635 pour Pms1 de levure. La taille importante de ces charnières est en bon accord avec les grands réarrangements observés par microscopie à force atomique sur la protéine MutL α de *S. cerevisiae* (Sacho et al., 2008).

Structure globale de la région C-terminale du complexe de Mlh1 et Pms1.
Une vue générale de la structure de la région C-terminale du complexe Mlh1-Pms1 est présentée sur les (Figure 3 - 20(A)&(B)). Les CTD de Mlh1 et Pms1 sont composés chacun de deux domaines. Ces sous-domaines seront appelés domaine de dimérisation et domaine régulateur par analogie avec l'appellation adoptée pour les MutL bactériens (Pillon et al., 2010a) (Figure 3 - 20(C)&(D)). Les domaines de dimérisation de Mlh1 et de Pms1 sont constitués des extrémités N-terminales et C-terminales des régions cristallisées. Ils sont constitués des acides aminés 526-569 et 697-769 de Mlh1 (en jaune clair dans la figure 3-21) et des acides aminés 651-711 et 815-873 de Pms1 (en bleu clair). Les domaines régulateurs de Mlh1 et Pms1 comprennent les zones centrales de la séquence de la région C-terminale. Ils sont constitués des acides aminés 510-525 et 570-697 de Mlh1 (en bleu) et 712-814 de Pms1 (en orange). Comme indiqué dans l'introduction, ce domaine a été appelé régulateur par les auteurs de la structure de MutL de *B. subtilis* (bsMutL) en référence à son rôle dans l'interaction de MutL avec le facteur de réplication de l'ADN appelé β clamp.

Structure de la région C-terminale de Mlh1. Le domaine de dimérisation du CTD de Mlh1 est composé de quatre hélices (α_0 , α_E , $\alpha_{E'}$ et α_F) et d'un feuillet β à 4 brins (β_1 , β_2 , β_3 et β_8). Les éléments de structures secondaires de Mlh1 et Pms1 ont été nommés par analogie avec ceux observés dans les structures de MutL bactériens. Les dénominations '0' ou '(prime)' désignent des éléments de structures secondaires absents chez les protéines bactériennes. La grande hélice α_E correspond à celle trouvée chez MutL. L'hélice α_0 n'est pas présente dans les structures de CTD de MutL. La seconde hélice $\alpha_{E'}$ correspond à une boucle dans les structures de MutL. L'hélice C-terminale α_F se détache du domaine de dimérisation et interagit essentiellement avec le CTD de Pms1. Cette dernière hélice contient le motif très conservé de l'extrémité C-terminale de Mlh1. Le feuillet β (brins β_1 , β_2 , β_3 et β_8) forme l'essentiel de l'interface de dimérisation.

Le domaine régulateur de Mlh1 est composé d'une région interne à la séquence du domaine C-terminal de Mlh1 (570-697) et d'une hélice située en position

N-terminale (509 à 524). Ce domaine est composé d'un feuillet β à 4 brins (brins β_4 , β_5 , β_6 et β_7) et de cinq hélices (α_0 , α_A , α_B , α_C et α_D). L'hélice α_0 n'est pas présente dans les structures du CTD de MutL. Les constructions utilisées pour résoudre les structures d'ecMutL et de bsMutL démarrent après cette hélice α_0 . Cette hélice α_0 , qui est située en séquence près du domaine de dimérisation, interagit en fait avec le domaine régulateur. Nous verrons plus loin que l'hélice α_0 participe à la structure du site S2 qui se trouve dans ce domaine régulateur (cf. §iii).

Structure de la région C-terminale de Pms1: Le domaine de dimérisation de Pms1 est composé de trois hélices (α_0 , α_0' et α_E) et d'un feuillet β à 4 brins (β_1 , β_2 , β_3 et β_8). Contrairement à ce que nous avons observé pour Mlh1, l'hélice additionnelle α_0 fait partie du domaine de dimérisation et non du domaine régulateur. Le domaine régulateur de Pms1 est constitué de deux hélices (α_B et α_C) et d'un feuillet β à 4 brins (β_4 , β_5 , β_6 et β_7). Ce dernier feuillet β est la région du complexe Mlh1-Pms1 la moins bien définie dans les cartes de densité électronique par rapport au reste de la structure. Le facteur d'agitation thermique moyen sur les atomes de la chaîne principale de cette région est de 132\AA^2 contre 75\AA^2 pour l'ensemble du complexe.

Structure des charnières entre les domaines de dimérisation et régulateur. Le passage entre les domaines de dimérisation et régulateur est effectué par l'hélice α_A et la boucle α_D - α_E de Mlh1. Pour Pms1, ce passage est effectué par les hélices α_A et α_D .

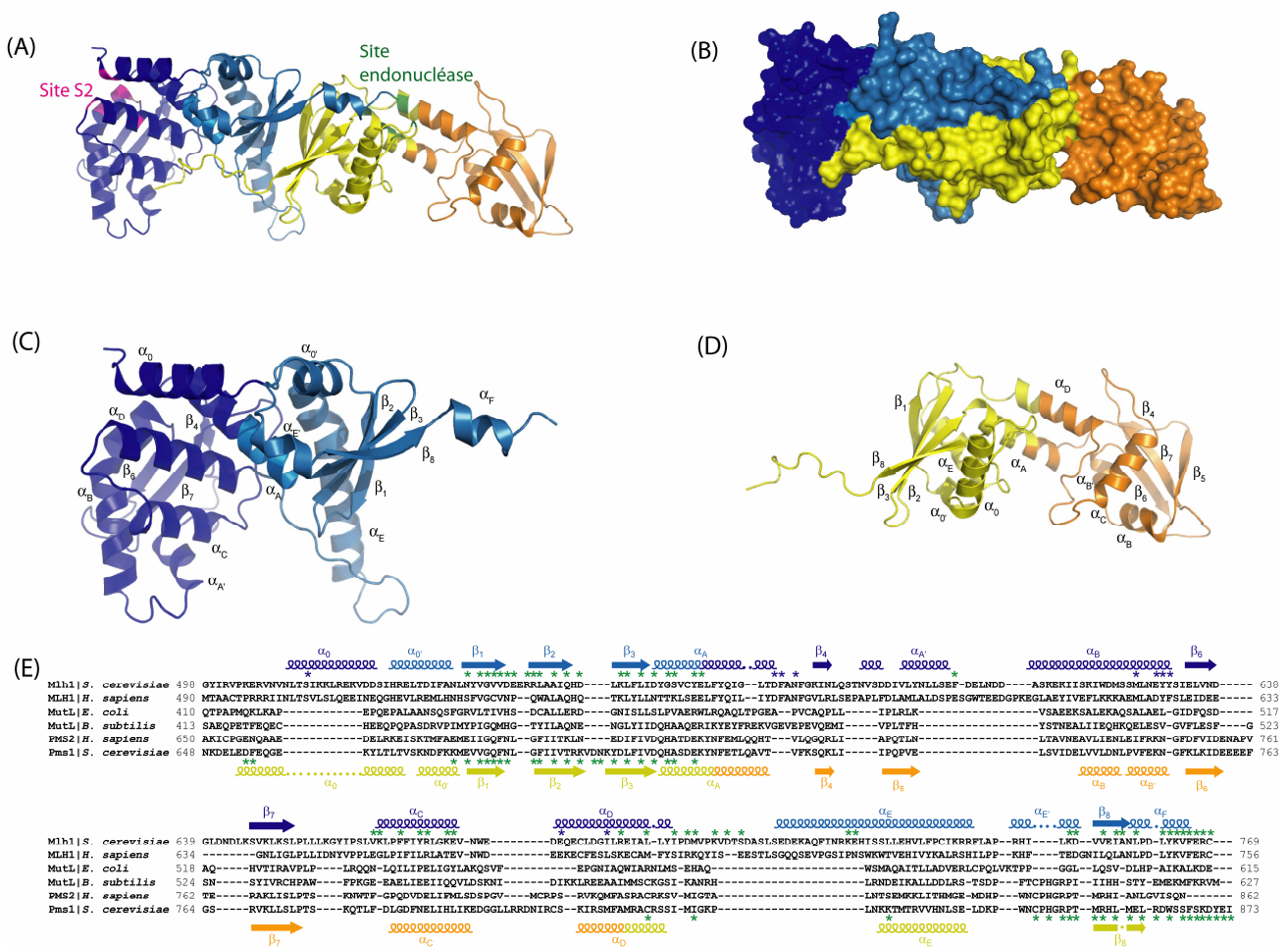


Figure 3 - 20 : Vue générale de la structure cristallographique du domaine C-terminal de Mlh1*Pms1 (A) Représentation en ruban et (B) en surface de la structure cristallographique ; le monomère de Mlh1 est représenté en bleu (bleu foncé et bleu clair pour, respectivement, les sous-domaines régulateur et de dimérisation) ; le monomère de Pms1 est représenté en jaune (jaune et orange pour, respectivement, les sous-domaines de dimérisation et régulateur) (C) Nomenclature des structures secondaires du monomère de Mlh1 (D) et de Pms1 (E) Alignement structural des CTD de Mlh1 chez *S. cerevisiae* et *H. sapiens*, de Pms1 chez *S. cerevisiae*, de PMS2 chez *H. sapiens* et de MutL chez *E. coli* et *B. subtilis* ; les domaines CTD humains sont alignés avec les domaines CTD de levure selon leur séquence par *K-align* ; les domaines CTD de levure et bactériens sont alignés selon leur structure cristallographique par *Superpose* ; les résidus impliqués dans le site S1 (hétérodimérisation) et dans le site S2 (motif MIP) sont surmontés d'une étoile verte et bleue, respectivement

Superposition des domaines de dimérisation des régions C-terminales de Mlh1, Pms1 et des MutL bactériens: La structure des monomères de Mlh1 et Pms1 a été superposée aux structures cristallographiques de MutL bactériens à l'aide du programme *Superpose* (Tableau 3 - 16) Dans un premier temps, les domaines de dimérisation et régulateur ont été superposés séparément aux domaines équivalents des protéines MutL. Le domaine de dimérisation de Pms1 se superpose très bien avec le domaine de dimérisation d'ecMutL avec une déviation (rmsd ou "root mean square deviation" ou racine de la moyenne du carré des déviations) de 1.8Å sur 53 C α . La superposition avec le domaine de dimérisation de bsMutL est encore meilleure (rmsd

de 1.5Å sur 63 C_α). (Tableau 3 - 16). Le domaine de dimérisation de Mlh1 présente des valeurs de rmsd plus grandes avec les régions équivalentes d'ecMutL et bsMutL (rmsd de 2.8Å sur 57 C_α et 3.1Å sur 60 C_α). Les valeurs supérieures de rmsd pour Mlh1 viennent d'une orientation différente de l'hélice α_E entre Mlh1 et les protéines MutL (Figure 3 - 21(A)).

Tableau 3 - 16 : Déviations entre les domaines C-terminaux d'ecMutL (Code PDB :1X9Z) et de bsMutL (Code PDB :3KDK) d'une part, et les domaines C-terminaux de Mlh1 et Pms1 d'autre part ; les rmsd (root mean square deviation) ainsi que le nombre de C_α superposés sont présentés. Les superpositions sont réalisées pour l'ensemble des domaines (en noir), le sous-domaine de dimérisation (en bleu clair pour Mlh1 ; en jaune pour Pms1) et le sous-domaine régulateur (en bleu foncé pour Mlh1 ; en orange pour Pms1) ; les superpositions et les valeurs de rmsd sont calculées à l'aide de l'option SSM (superposition des structures secondaires) du programme Superpose

	CTD de Mlh1	CTD de Pms1
CTD d'ecMutL	Global: 3.2Å (155)	Global: 3.1Å (95)
	Dim: 2.8Å (57)	Dim: 1.8Å (53)
	Reg: 2.3Å (83)	Reg: 2.3Å (71)
CTD de bsMutL	Global: 4.4Å (142)	Global: 1.6Å (99)
	Dim: 3.1Å (60)	Dim: 1.5Å (63)
	Reg: 2.9Å (85)	Reg: 2.8Å (76)

Superposition des domaines régulateurs des régions C-terminales de Mlh1, Pms1 et des MutL bactériens: Les domaines régulateurs de Mlh1 et Pms1 ont été ensuite superposés aux domaines équivalents de MutL. Le domaine régulateur de Pms1 se superpose bien avec ceux de MutL bien que les valeurs de rmsd soient supérieures à celles obtenues avec le domaine de dimérisation. La principale différence est observée au niveau d'une boucle entre le brin β_7 et l'hélice α_C ainsi que d'une orientation différente de l'hélice α_D . Le domaine régulateur de Mlh1 se superpose avec le domaine régulateur des protéines MutL avec des valeurs similaires à celles obtenues pour Pms1. Mlh1 présente entre le brin β_4 et l'hélice α_B une hélice, l'hélice α_A , alors que les protéines MutL ont un brin β_4 plus long.

Superposition des régions C-terminales entières des régions C-terminales de Mlh1, Pms1 et des MutL bactériens. La Figure 3 - 21(A) présente les régions C-terminales entières de Mlh1 et MutL, après superposition des domaines de dimérisation. Comme on peut le voir sur cette figure, le feuillet β du domaine de dimérisation de Mlh1 et de bsMutL se superposent bien ainsi que les hélices α_A , α_C et α_D .

La Figure 3 - 21(B) présente la superposition des régions C-terminales entières de Pms1 et MutL, après superposition des domaines de dimérisation. Comme on peut le voir sur cette figure, le feuillet β du domaine de dimérisation de Mlh1 et celui de bsMutL se superposent bien ainsi que les hélices α_A , α_D et α_E . Le domaine régulateur

de Pms1 occupe une position très différente de celui de MutL. La position de ce domaine a subi une rotation de 70° autour des charnières entre les domaines de dimérisation et régulateur. Cette rotation a comme points d'articulation les acides aminés 718 à 721 et 767 à 805 de Pms1.

Cette observation soulève un certain nombre de questions. Le mouvement du domaine régulateur de Pms1 correspond-il à une propriété intrinsèque de Pms1 ? Ce mouvement est-il induit ou favorisé par l'empilement cristallin ? Les autres régions C-terminales de MutL bactériens ou eucaryotes comme Mlh1 sont-elles susceptibles de présenter le même type de mouvement ? Nous ne disposons pas aujourd'hui d'éléments suffisants pour répondre complètement à cette question. La comparaison des structures de MutL(CTD) et Pms1 (CTD) montre une re-orientation d'environ 70°. L'analyse des modes normaux de la structure d'ecMutL effectuée lors de la préparation de modèles pour le remplacement moléculaire, propose des mouvements de rotation entre les sous-domaines de dimérisation et régulateur. L'amplitude des mouvements proposés (environ 15°) est également plus faible que celle observée entre la structure d'ecMutL et celle de Pms1. La détermination de la structure d'un second complexe comme la région C-terminale de MLH1*PMS2 humain ou de Mlh1*Pms1 de levure dans une autre forme cristalline permettra probablement d'évaluer l'impact de l'empilement cristallin sur ces mouvements entre domaine de dimérisation et régulateur.

On peut noter que le réarrangement majeur du domaine régulateur de Pms1 entraîne une structure globale différente de Pms1 par rapport à la région C-terminale de MutL. Cette différence explique en partie les difficultés rencontrées lors du remplacement moléculaire avec la structure de la région C-terminale de MutL comme modèle.

En conclusion, les alignements structuraux indiquent une conservation globale du repliement entre les régions C-terminales de Mlh1 et Pms1. Cette similarité structurale vient compléter celle observée sur les régions N-terminales de Mlh1 et Pms1 ainsi que sur les protéines MutS. La structure cristallographique était nécessaire pour établir cette similarité étant donné le peu de conservation entre les séquences de Mlh1 et Pms1 avec les MutL bactériens (9% d'identité de séquence entre les domaines C-terminaux de Mlh1 et d'ecMutL; 12% entre ceux de Pms1 et d'ecMutL) (**Figure 3 - 20(E)**).

Analyse de la qualité des prédictions de structure secondaire de Mlh1 et Pms1 à la lueur de la structure cristallographique. Nous avons ré-analysé à la lueur de la structure cristallographique du C-ter de Mlh1-Pms1, les résultats de prédiction

de structures secondaires obtenues sur ces deux protéines³². Les éléments de structures secondaires de Mlh1 sont tous prédits, à l'exception de l'hélice α_F à l'extrémité C-terminale. Ces éléments sont prédits avec une précision de plus ou moins deux acides aminés. Les programmes prédisent bien la présence de l'hélice α_0 . Cependant sa position dans la séquence suggérerait éventuellement une interaction de cette hélice avec le domaine de dimérisation. Les éléments de structure secondaire de Pms1 sont également bien prédits. La principale différence par rapport aux prédictions concerne l'extrémité C-terminale de Pms1 (860 à 873). Cette région est prédite en hélice. Elle forme en fait une région étendue de 27Å de longueur en interaction avec Mlh1.

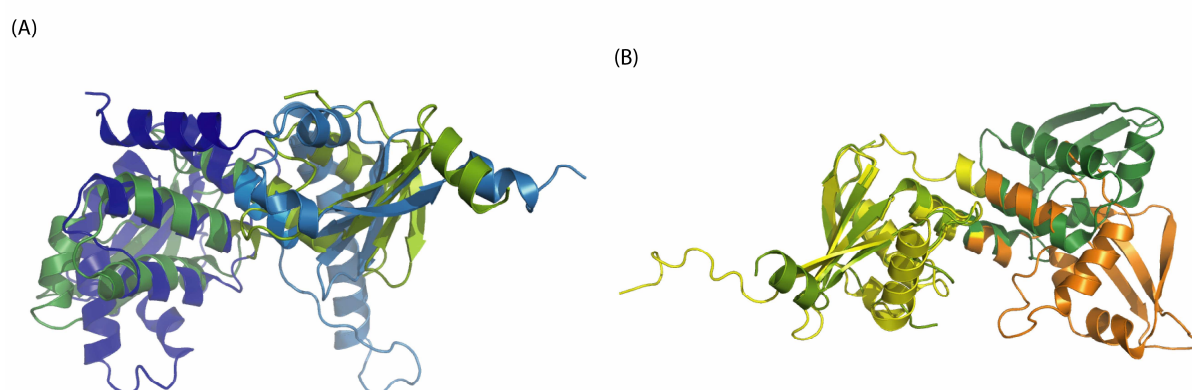


Figure 3 - 21 : Superposition des domaines C-terminaux de bsMutL et de MutL α (A) Superposition des CTD de Mlh1 (en bleu) et de bsMutL (en vert) (B) Superposition des CTD de Pms1 (en jaune) et de bsMutL (en vert) via leur sous-domaine de dimérisation (résidus 651-711 & 815-873 de Pms1 et 433-471 & 571-624 de bsMutL)

ii. Analyse du site d'hétérodimérisation entre Mlh1 et Pms1

L'interface d'hétérodimérisation implique 53 résidus du domaine C-terminal de Mlh1³³ et 46 résidus du domaine C-terminal de Pms1³⁴. Ces acides aminés enfouissent lors de la formation du complexe Mlh1-Pms1 une surface de 1790Å² pour Mlh1 et 1820Å² pour Pms1 (Figure 3 - 22). La surface enfouie totale de ce complexe est de 3610Å². Cette valeur se situe dans la région supérieure des valeurs de surfaces enfouies observées dans les structures cristallographiques de complexes protéine-protéine (surface enfouie moyenne de 1600Å² \pm 400Å² (Lo Conte et al., 1999). En comparaison, les surfaces enfouies dans les dimères de bsMutL et d'ecMutL sont près de deux fois plus petites (surfaces enfouies de 2060Å² pour bsMutL et de 1770Å² pour ecMutL).

³² La comparaison de structures secondaires prédites et réelles est présentée en ANNEXE 13

³³ L'ensemble des résidus de Mlh1 impliqués dans l'interface de dimérisation est indiqué en ANNEXE 14

³⁴ L'ensemble des résidus de Pms1 impliqués dans l'interface de dimérisation est indiqué en ANNEXE 15

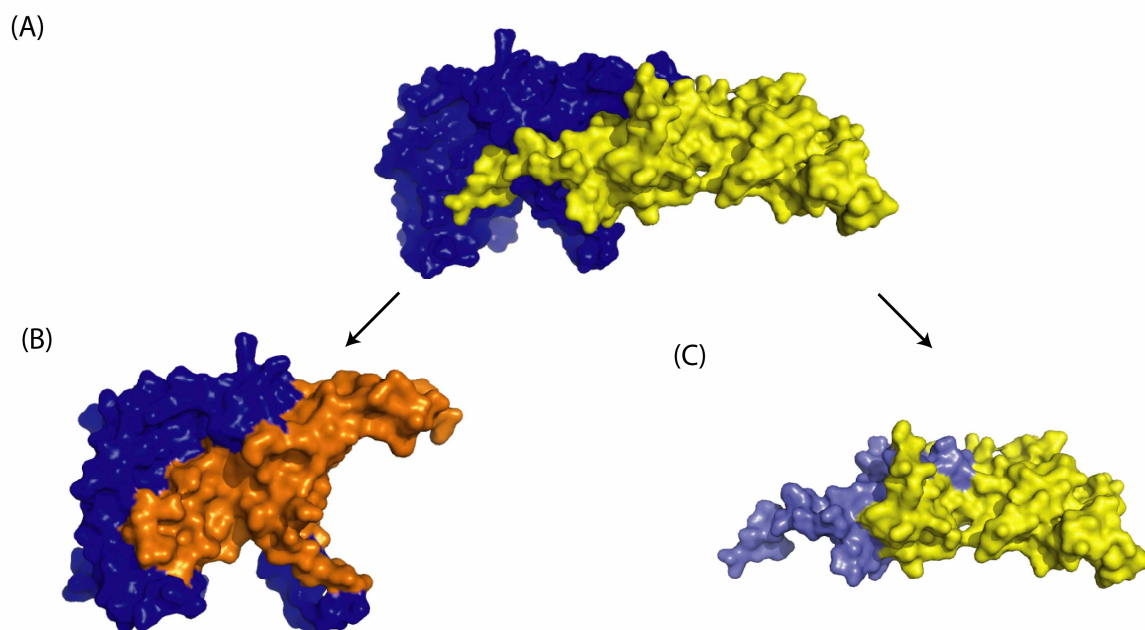


Figure 3 - 22 : Analyse de la surface de dimérisation du domaine C-terminal de MutL α (A) Représentation en surface du CTD de Mlh1*Pms1 ; le CTD de Mlh1 est représenté en bleu, le CTD de Pms1 est représenté en jaune (B) Représentation en surface du CTD de Mlh1 seul ; les résidus impliqués dans l'interface de dimérisation sont représentés en orange (C) Représentation en surface du CTD de Pms1 seul ; les résidus impliqués dans l'interface de dimérisation sont représentés en bleu clair

L'interface d'hétérodimérisation met en jeu trois zones principales:

(i) La première zone est constituée des feuillets β des domaines de dimérisation de Mlh1 et Pms1. Ces feuillets constituent une part importante du site d'hétérodimérisation (1418\AA^2 sur 3610\AA^2 soit 40%). L'interaction est effectuée par 9 résidus hydrophobes de Mlh1 (V539, V542, L548, A550, L557 et L559, V754, I756 et A757) et 9 résidus hydrophobes de Pms1 (V677, F680, F684, I686, L697, I699, L858, M859 et L861) (**Figure 3 - 23(A)**). En complément de ce cœur hydrophobe, quatre ponts salins viennent stabiliser l'interface en périphérie du feuillet β (K724^M-E675^P, K673^P-D706^M, K690^P-E544^M, K829^P-E725^M) (**Tableau 3 - 17**). La superposition des régions de dimérisation des structures de MutL(CTD) d'*E. coli*, de *B. subtilis* et de *N. gonorrhoeae* sur Mlh1(CTD) et Pms1(CTD) montrent que cette première zone du site d'hétérodimérisation entre Mlh1 et Pms1 est similaire d'un point de vue structural au site d'homodimérisation des protéines MutL bactériennes (**Figure 3 - 23(E)**).

Tableau 3 - 17 : Ponts salins présents à l'interface entre le CTD de Mlh1 et le CTD de Pms1

Résidu basique	Résidu Acide	Distance (Å)
N _{η1} R547 ^M	O _{δ1} D870 ^P	3.66
N _{η2} R547 ^M	O _{δ1} D870 ^P	3.16
N _{η2} R547 ^M	O _{δ2} D870 ^P	2.57
N _ζ K724 ^M	O _{ε1} E675 ^P	2.84
N _ζ K673 ^P	O _{δ1} D706 ^M	3.28
N _ζ K673 ^P	O _{δ2} D706 ^M	2.89
N _ζ K690 ^P	O _{ε1} E544 ^M	3.66
N _ζ K829 ^P	O _{ε1} E725 ^M	3.79

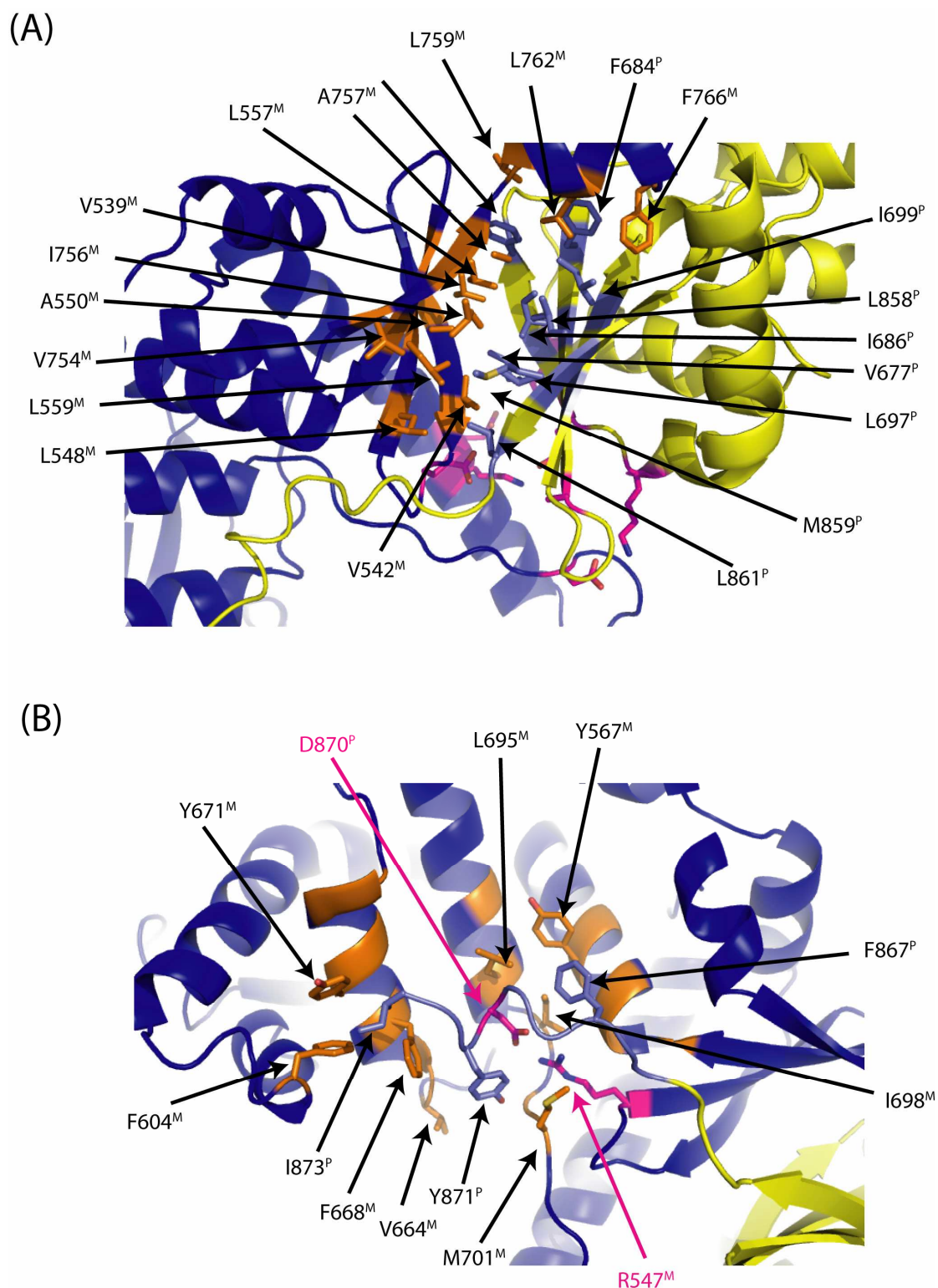


Figure 3 - 23 : Nature des interactions à l'interface d'hétérodimérisation (A) Interaction au niveau des feuillets β des sous-domaines de dimérisation de Mlh1 et Pms1 ; les résidus hydrophobes, impliqués dans l'interface de Mlh1 et Pms1 sont représentés en orange et bleu clair, respectivement ; les résidus impliqués dans un pont salin sont représentés en magenta (B) Interaction au niveau de l'extension C-terminale de Pms1 ; les résidus impliqués dans l'interface de Mlh1 et Pms1 sont représentés en orange et bleu clair, respectivement et les chaînes latérales des résidus hydrophobes sont explicitées ; les résidus impliqués dans un pont salin sont représentés en magenta et leurs chaînes latérales est représentés en bâton

(ii) La seconde zone du site d'hétérodimérisation est constituée par l'extrémité C-terminale de Mlh1. Cette extrémité forme une excroissance qui vient embrasser le domaine de dimérisation de Pms1. Elle est composée des 8 derniers acides aminés, ${}_{762}\text{LYKVFERC}_{769}$, de Mlh1 qui sont très conservés à travers l'ensemble des séquences eucaryotes. Les cinq premiers résidus font partie de l'hélice α_F . Trois de ces résidus (L759, L762 et F766) sont hydrophobes et sont au contact de résidus hydrophobes de Pms1 (F680, F684 et I699) **Figure 3 - 23(A)**. L'hélice α_F possède une orientation différente de l'hélice équivalente présente chez les protéines MutL bactériennes (**Figure 3 - 23(A)**). Les trois derniers résidus du motif conservé ${}_{762}\text{LYKVFERC}_{769}$ forment une structure étendue. Les résidus E767 et R768 interagissent avec Pms1 par les atomes de la chaîne principale. Le dernier résidu C769 se retrouve alors à proximité du site endonucléase de Pms1 (voir la description du site endonucléase §iv.). La surface enfouie de Mlh1 et Pms1 dans la zone de l'extrémité C-terminale de Mlh1 est de 722\AA^2 sur 3610\AA^2 soit 20% de la surface enfouie totale.

(iii) La troisième et dernière zone du site d'hétérodimérisation est constituée par l'extrémité C-terminale de Pms1. Cette région de Pms1 adopte, de façon inattendue, une conformation étendue sur 13 acides aminés et une longueur de 27\AA . La densité électronique définit sans ambiguïté la trace de cette région de Pms1. Contrairement aux MutL bactériens et au monomère Mlh1, Pms1 ne possède pas de structure en hélice dans cette zone. L'extrémité C-terminale de Pms1 vient embrasser la région de dimérisation de Mlh1. Elle contribue pour 1552\AA^2 à la surface d'hétérodimérisation soit 43%. L'extrémité C-terminale de Pms1 interagit avec la première moitié de l'hélice α_A (résidus 562 à 568), une partie de l'hélice α_E (K751 et D752) et l'insertion 1 de Mlh1 (résidus 698 à 708). De par sa structure étendue, cette extrémité C-terminale interagit également avec des acides aminés du domaine régulateur de Mlh1 (résidus 664 à 676 de l'hélice α_C et la deuxième moitié de l'hélice α_D (résidus 691 à 697) de Mlh1). L'extension C-terminale de Pms1 possède trois résidus hydrophobes (F867, Y871 et le dernier acide aminé de la séquence de Pms1, I873). Ces acides aminés sont en contact avec Mlh1: F867^P est en contact avec Y567^M, L695^M et I698^M; Y871^P et I873^P sont en contact avec F604^M, V664^M, F668^M et Y671^M. L'interaction de l'extrémité N-terminale de Pms1 est également stabilisée par un pont salin entre les résidus D870^P et R547^M (**Figure 3 - 23(B)**).

La caractérisation du site d'hétérodimérisation de Mlh1*Pms1 permet de ré-analyser les données de mutagenèse dirigée et double-hybride reportées par différents groupes. Dans notre étude précédente (Dherin et al., 2009) et dans une étude du laboratoire de R. Liskay (Pang et al., 1997), des délétions progressives de la région C-terminale de Mlh1 ont montré que la délétion des dix derniers acides aminés, Mlh1 Δ C10 n'affecte pas l'interaction avec Pms1 mesuré par double hybride. De façon intéressante, un plasmide contenant le mutant Mlh1 Δ C10 complémente une souche

mlh1Δ de façon moins efficace que le plasmide contenant la protéine Mlh1 sauvage. A la lueur de notre structure, ces résultats suggèrent que la dernière hélice α_F de Mlh1 bien que très conservée chez les eucaryotes n'est pas indispensable pour l'interaction avec Pms1. Ceci suggère également que cette extrémité C-terminale de Mlh1 joue un rôle additionnel comme cela sera discuté un peu plus loin. La délétion des 21 derniers acides aminés de Mlh1, Mlh1 Δ C21, entraîne une perte d'interaction avec Pms1 mais également avec Exo1 et Ntg2. Les résultats négatifs obtenus avec tous les partenaires pour Mlh1 Δ C21 s'expliquent bien avec la structure cristallographique car cette construction correspond à la délétion d'éléments de structures secondaires importants pour le repliement de la protéine (l'hélice α_F , le brin β_8 et la moitié de l'hélice α_E). Le laboratoire de R. Liskay a également montré (Pang et al., 1997) que la délétion des 13 derniers acides aminés de Pms1 détruisait l'interaction avec Mlh1. Avec la structure du complexe, ce résultat suggère que l'extrémité étendue de Pms1 est essentielle pour l'interaction avec Mlh1.

Différents mutants ponctuels, simple ou multiple, de ce site d'hétérodimérisation ont également été étudiés. Un crible génétique a été réalisé par le laboratoire de R. Kolodner pour identifier des mutations qui augmentent significativement le faible taux mutateur des souches *exo1Δ*. Ce crible a permis d'isoler un mutant *mlh1-R547K* (Amin et al., 2001). Cette mutation s'accompagne seule d'un phénotype mutateur important sur un essai de réversion de décalage de lecture *Lys2* dans une séquence répétée A_{10} (augmentation du taux de mutation 866 fois par rapport à une souche sauvage). Le double mutant *exo1Δ mlh1-R547K* présente un phénotype mutateur plus fort en accord avec le crible génétique effectué (augmentation du taux de mutation d'un facteur 7 600 contre 18 900 pour une souche *msh2Δ*). Dans notre étude du site S2 (Dherin et al., 2009), la mutation Mlh1-R547A a été analysée en double hybride. Ce mutant ne présente plus d'interaction avec Pms1 et conserve une interaction avec Exo1 et Ntg2. Ces deux résultats suggèrent que les simples mutations Mlh1-R547K ou Mlh1-R547A, qui se situent au cœur de l'interaction entre l'extrémité C-terminale de Pms1 et Mlh1, déstabilisent l'ensemble des interactions dans cette région du site d'hétérodimérisation et par là même l'interaction Mlh1-Pms1. Une troisième étude réalisée dans le laboratoire d'E. Alani vient souligner le rôle des interactions entre Mlh1 et Pms1 autour de la position R547^M. Ce laboratoire a ainsi étudié 60 mutants sur la protéine Mlh1 entière, chaque mutant portant des substitutions sur deux à cinq acides aminés chargés et proches en séquence (Argueso et al., 2003). Ces 60 mutants ont été analysés à l'aide de tests semi-quantitatifs de mutations et de tests sur l'efficacité des crossover en méiose. Dans la région C-terminale, le motif ${}_{543}\text{DEERR}_{547}$ de Mlh1 a été muté en alanine. Ce pentuple mutant n'est plus fonctionnel dans les deux tests de mutation regardés (un essai semi-quantitatif de résistance à la canavanine et une mesure du taux de réversion *hom3-10*) (Argueso et al., 2003). Ce mutant, appelé *mlh1-2* par les auteurs, n'interagit plus avec Pms1 et Mlh3 en double hybride. Il présente un niveau presque

sauvage de viabilité des spores et de formation des crossover. La position R547 fait partie de ce pentuple mutant. La déstabilisation de l'interaction avec Pms1 et les phénotypes mutateurs sont en accord avec les données sur les mutants ponctuels R547K et R546A. De façon intrigante, ce pentuple mutant n'interagit plus en double hybride avec Mlh3 tout en conservant un niveau presque sauvage de recombinaison méiotique. On peut noter qu'avant la détermination de la structure du complexe Mlh1*Pms1, la modélisation moléculaire ne permettait pas de rendre compte de l'effet des mutations autour de la position R547 car aucun modèle ne prédisait le caractère étendue de la structure de l'extrémité C-terminale de Pms1.

Dans une étude combinant la modification de surface des protéines et la spectrométrie de masse, le laboratoire de K. Tomer et T. Kunkel a identifié trois positions de Mlh1, K665, K675 et K704, qui semblent exposées sur la protéine seule mais enfouies lors de la formation du complexe avec Mlh1. En effet, ces trois lysines de Mlh1 ne peuvent plus être acétylées, en présence d'acide acétique anhydride, lorsque le dimère Mlh1*Pms1 est formé alors qu'elles sont acylées en l'absence de Pms1 (Cutalo et al., 2006). Ces données sont en accord avec notre structure 3D. Les deux premières lysines, K665 et K704, sont situées sur l'hélice α_C à proximité de l'extension C-terminale de Pms1. La troisième lysine fait partie de la boucle $\alpha_D - \alpha_E$ située à proximité du feuillet β de Pms1 impliqué dans l'hétérodimérisation. En absence de structure 3D, les auteurs de cette dernière étude avaient proposé un modèle alternatif de dimérisation de Mlh1*Pms1 radicalement différent de celui observé dans la structure présentée ici.

En conclusion, les régions C-terminales de Mlh1 et Pms1 forment un hétérodimère par des interactions entre les feuillets β de leur domaine de dimérisation avec un positionnement de ces deux feuillets similaire à celui observé chez les protéines MutL(CTD) bactériennes. En revanche, la structure cristallographique de notre complexe met en évidence un rôle original des extrémités C-terminales de Mlh1 et Pms1. La structure de la surface d'hétérodimérisation permet de comprendre un nombre important de données biochimiques et génétiques que les travaux de modélisation moléculaire ne permettaient pas d'expliquer.

iii. Etude structurale du site S2 d'interaction avec Ntg2, Exo1 et Sgs1

Nous avons déterminé au cours de la première partie de cette thèse, 10 positions de Mlh1 importantes pour l'interaction avec Ntg2, Exo1 ou Sgs1 en combinant tests de double-hybride et modélisation moléculaire. A présent, il est donc intéressant d'analyser la position de ces résidus et de comparer la structure cristallographique avec le modèle utilisé pour guider la mutagenèse dirigée.

La structure cristallographique du domaine C-terminal de MutL α montre que 9 positions identifiées à partir du modèle par homologie, sont situées dans la même région du domaine régulateur de Mlh1 (**Figure 3 - 24(A)**) (la dixième position

correspond à la sérine S513 qui n'est pas présente dans le modèle). Les résidus D577, F578 et N580 se trouvent au bout de l'hélice α_A tandis que les résidus M626, E629, Y630 et Y631 sont situés sur l'hélice α_B . Enfin les résidus E682 et I689 se trouvent sur l'hélice α_D . Sur ces 9 positions, 6 sont accessibles au solvant. Il s'agit des positions D577, N580, M626, E629, Y630 et E682. Ainsi, notre structure montre que, concernant le site S2, le modèle construit par homologie, est en bon accord avec structure 3D résolue.

La dixième position, la sérine S513, n'est pas présente dans les modèles construits par homologie car les constructions utilisées pour la cristallisation des CTD de MutL bactériens démarrent après cette position. Cette position est par contre présente dans la construction de Mlh1 (485-769) utilisée dans notre étude. La sérine S513 se trouve sur la première hélice modélisée, l'hélice α_0 , à l'extrémité N-terminale de Mlh1. De façon intéressante, l'hélice α_0 est positionnée près du domaine régulateur et la sérine S513 est localisée à proximité des 9 autres résidus du site S2 (**Figure 3 - 24A**). On peut noter que la modélisation ne permettait pas de prédire son rôle. Cette position avait été testée en double hybride car elle correspondait à un acide aminé polaire conservé dans un alignement multiple de séquences Mlh1 eucaryotes.

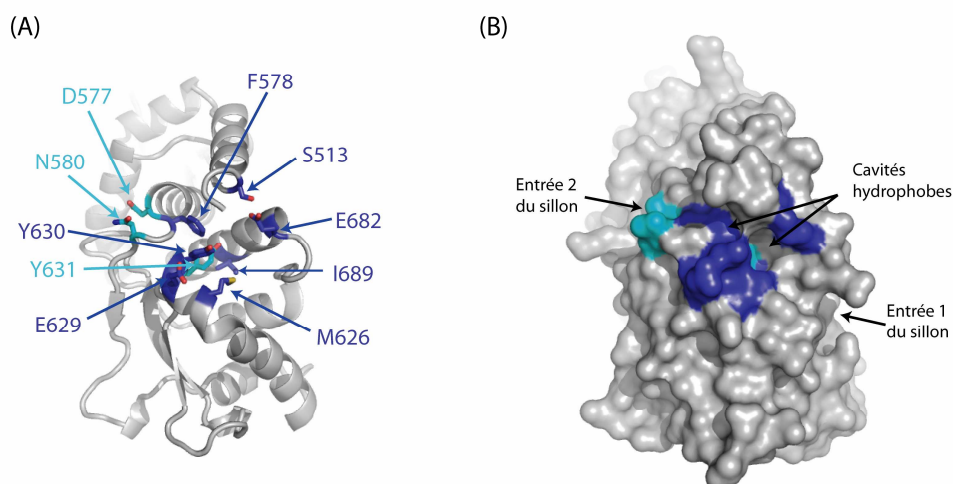


Figure 3 - 24: Structure du site S2 du domaine C-terminal de Mlh1 (A) Localisation des 10 positions identifiées du site S2; les résidus (-) en double-hybride sont représentés en bleu, les résidus (+/-) sont représentés en cyan (B) Représentation en surface du site S2 à la surface du CTD de Mlh1; le code de couleur est identique à la figure A

La représentation en surface de la région C-terminale de Mlh1 (**Figure 3 - 24(B)**) montre que le site S2 forme une région contigüe et présente une forme de sillon. Le centre du sillon est composé de deux acides aminés hydrophobes, M626 et I689, et de trois aromatiques, F578, Y630 et Y631. Ces 5 résidus forment deux cavités hydrophobes. Les deux glutamates, E629 et E682, du site S2 sont situés à l'entrée de ces deux cavités. L'acide aminé E682 fait une liaison hydrogène intramoléculaire par sa chaîne latérale avec le groupe hydroxyle de la sérine S513 (distance ($O_{\epsilon 2}^{E682} - O_{\gamma}^{S513}$) = 2.7 Å). Enfin les résidus D577 et N580 se trouvent à la périphérie du site S2.

La structure cristallographique du CTD de Mlh1 montre que les acides aminés hydrophobes et aromatiques identifiés comme importants pour l'interaction avec les protéines contenant un motif MIP définissent deux poches. Nous avons dans un premier temps décidé de modéliser l'interaction d'un peptide pExo1 de 16 acides aminés ($_{439}\text{TLKDTRSKFFENKPSMT}$) avec le site S2. Ce peptide présente une constante de dissociation de $5.6\mu\text{M}$ avec la région C-terminale de Mlh1. La modélisation a été réalisée par B. Gilquin au laboratoire par dynamique moléculaire à quatre dimensions sous contraintes ambiguës. Une contrainte ambiguë impose de positionner les deux phénylalanines, F447^E et F448^E, du motif MIP d'Exo1 à proximité des cinq résidus hydrophobes du site S2 (F578^M, M626^M, Y630^M, Y631^M et I689^M). En revanche, aucune contrainte n'est appliquée dans les premières étapes de la dynamique moléculaire sur la sérine, S445^E, du motif MIP d'Exo1. Lors du refroidissement de dynamique moléculaire, une contrainte ambiguë est imposée sur le groupe hydroxyle de la sérine S445^E afin d'orienter son hydroxyle vers un atome polaire de Mlh1. Enfin nous avons choisi d'imposer quatre contraintes ambiguës entre les résidus hydrophobes V521^M, L575^M, I616^M et W679^M de Mlh1 et les résidus du peptide Exo1 situés de part et d'autre du motif MIP. Ces contraintes orientent la position du peptide pour qu'il épouse la forme du sillon observé sur Mlh1 de part et d'autre des deux poches hydrophobes.

Nous avons généré 2 200 structures et analysé les 50 structures qui présentent les plus faibles énergies de contrainte. Ces structures peuvent être classées en cinq groupes. Ces configurations se distinguent par la nature des acides aminés du motif MIP qui occupent les deux cavités hydrophobes du site S2 et par la position de la sérine du motif MIP. (i) Dans un premier groupe de 13 structures, le peptide adopte une conformation qui ne suit pas la forme du sillon. (ii) Dans un second groupe de 19 structures, le peptide Exo1 adopte une conformation où chaque phénylalanine du motif MIP est enfouie dans une des deux cavités hydrophobes et la sérine forme une liaison hydrogène avec l'atome d'oxygène du carbonyle du résidu E680^M (**Figure 3 - 25(A)**). (iii) Dans un troisième groupe de 12 structures, le peptide adopte une conformation où seule la première phénylalanine, F447^E, du motif MIP de Exo1 s'enfouit dans une cavité hydrophobe et la sérine forme une liaison hydrogène avec l'hydroxyle de la tyrosine Y630^M (**Figure 3 - 25(B)**). (iv) Dans un quatrième groupe de 8 structures, le peptide adopte une conformation où seule la seconde phénylalanine, F448^E, du motif MIP de Exo1 est enfouie dans une cavité hydrophobe et la sérine forme une liaison hydrogène avec la fonction acide du glutamate E629 (**Figure 3 - 25(C)**). (v) Enfin, dans le cinquième groupe de 6 structures, le peptide adopte une conformation où chaque phénylalanine du motif MIP est enfouie dans une des deux cavités hydrophobes et la sérine forme un contact polaire avec la fonction acide du glutamate E629 (**Figure 3 - 25(D)**).

Le dernier groupe, bien que d'occurrence plus faible, présente les caractéristiques les plus intéressantes par rapport aux données fonctionnelles et aux

mesures de calorimétrie. Les deux phénylalanines, F447^E et F448^E, du motif MIP sont enfouies dans les cavités hydrophobes. Cette conformation est en bon accord avec la perte d'interaction observée en double hybride lors des simples mutations F447A ou F448A sur Exo1 (Gellon et al., 2002). La liaison hydrogène proposée dans ce dernier groupe de modèles entre la fonction hydroxyle de la sérine S445^E et la chaîne latérale du glutamate E629^M est également en bon accord avec la perte d'interaction observée pour les deux mutants Exo1-S445A ou Mlh1-E629A. Dans les différents modèles d'ancrage obtenus, nous observons que la sérine S513 n'a pas de rôle direct dans l'interaction. La perte d'interaction observée avec sa mutation semble provenir d'un rôle structural et plus particulièrement du positionnement au travers d'une liaison hydrogène de la position E682, cette position participant de façon importante à la forme du sillon et également à la fixation du peptide (distance ($O_{\epsilon 1}^{E682} - O^{F448}$)=2.8Å). Du côté du peptide, ce modèle confirme les résultats obtenus en calorimétrie pour les positions R444^E et K446^E d'Exo1. Les chaînes latérales de ces deux acides aminés sont dirigées vers le solvant dans le modèle expliquant leur moindre contribution observée en double-hybride (Gellon et al., 2002) et en calorimétrie. Enfin, nous avons montré pour Exo1, qu'un peptide de 8 acides aminés présente la même affinité qu'un peptide de 12 ou 16 acides aminés. Ces données suggèrent que les interactions modélisées de part et d'autre de la séquence, ₄₄₃TRSKFFNK₄₅₀ ne contribuent que faiblement à l'affinité observée.

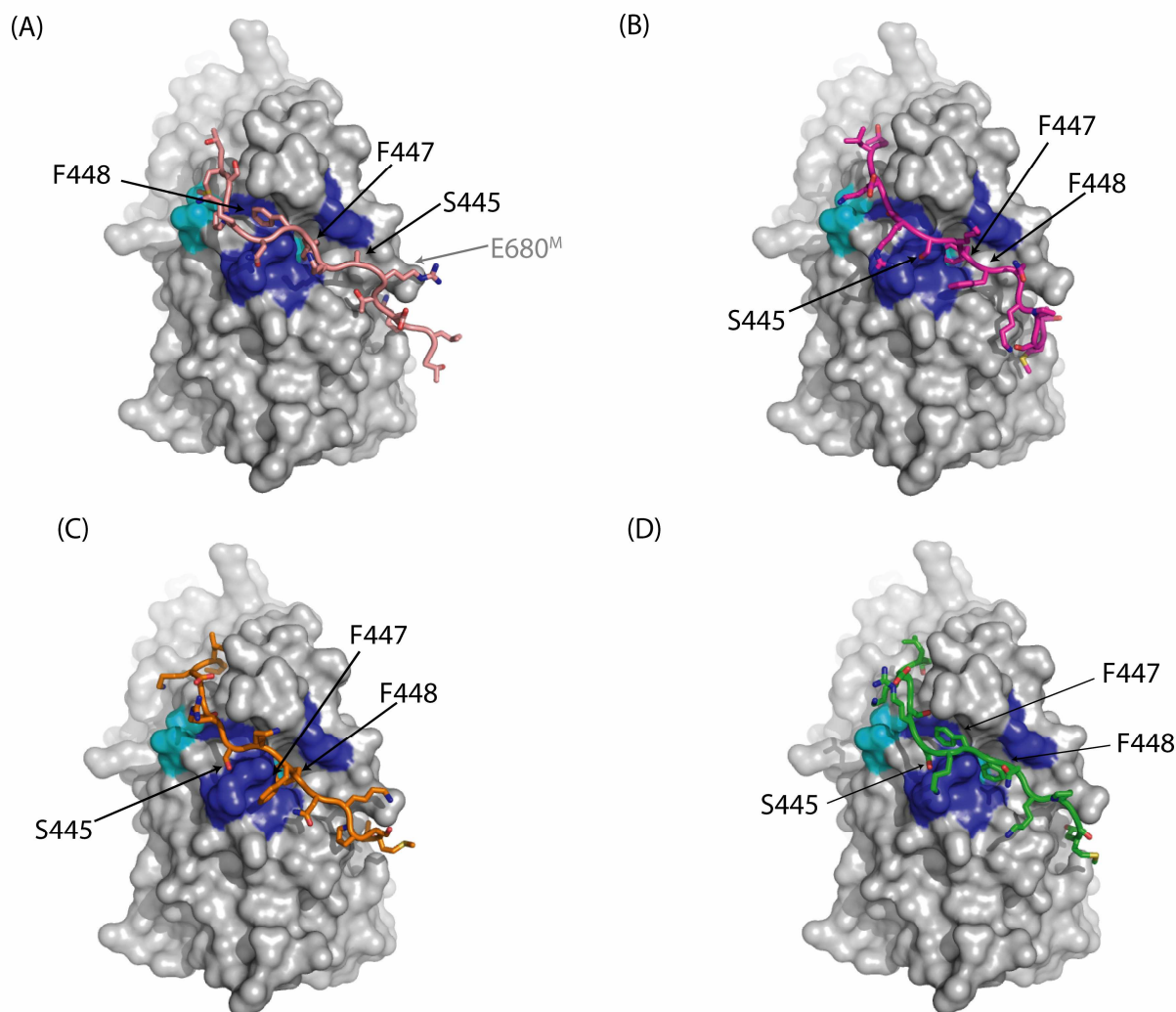


Figure 3 - 25: Modélisation de l'ancrage du peptide pExo1(16-mère) sur le site S2 de Mlh1. Les quatre configurations principales obtenues par dynamique moléculaire sont présentées; le code de couleur de la surface de Mlh1 est identique à celui utilisé précédemment; le peptide est représenté en (A) saumon (B) rose (C) orange (D) vert. La configuration (D) nous semble être celle qui est le plus en accord avec les données d'interaction Mlh1-peptide Exo1.

En conclusion, la structure cristallographique de Mlh1*Pms1 présentée dans cette étude valide la caractérisation du site S2 de Mlh1 proposée initialement par des approches de double-hybride et de modélisation moléculaire (Gellon et al., 2002). La structure cristallographique définit la structure du site S2. On observe, dans ce site S2, la présence de deux poches hydrophobes et d'un sillon susceptibles d'accueillir le motif MIP des protéines Exo1, Ntg2 et Sgs1. La modélisation par dynamique moléculaire du complexe entre Mlh1 et un peptide Exo1 permet de proposer différents modes d'ancrage de ce peptide. Un objectif pour la suite est de déterminer la structure cristallographique d'un complexe ternaire entre la région C-terminale du complexe Mlh1*Pms1 et un peptide contenant un motif MIP pour mieux comprendre les interactions autour du site S2.

iv. Etude structurale du site endonucléase de Pms1

Identification de l'ion métallique lié au site de Pms1 dans le cristal

Les structures cristallographiques des CTD de MutL bactériens contenant un site endonucléase et les analyses bioinformatiques des séquences Pms1 eucaryotes ont permis d'identifier trois motifs conservés de Pms1 susceptibles de constituer le site endonucléase. Ces observations ont été confirmées par des travaux de mutagenèse effectués par le laboratoire de P. Modrich (Kadyrov et al., 2006). Ces trois motifs sont: $_{701}\text{DQHASDEKYNFE}$ sur l'hélice α_A , $_{816}\text{ACR}$ sur l'hélice α_D et $_{848}\text{CPHGRP}$ qui forme un coude β entre l'hélice α_E et le brin β_8 . L'analyse de la position de ces motifs sur la structure du CTD de Mlh1*Pms1 montre que ces trois motifs forment un site contigu.

Nous avons identifié, au cours du phasage du complexe Mlh1-Pms1, deux diffuseurs anomaux au cœur de ce site, comme évoqué dans la partie des résultats relative à la détermination de structure. Ces deux diffuseurs anomaux, forment en fait un seul et même pic anomal de forme ellipsoïdale. On peut recenser les cations et anions métalliques susceptibles de présenter un signal anomal et utilisés au cours de l'expression, la purification et la cristallogenèse du complexe Mlh1*Pms1. L'expression du complexe Mlh1*Pms1 natif est réalisée en milieu LB qui contient des extraits de levure et donc toute sorte d'ions métalliques associés aux protéines de levure en quantité physiologique. Nous n'avons ajouté aucun ion métallique à cette étape. Lors de la purification, aucun ion métallique n'a été introduit dans les tampons hormis lors de l'étape de lyse cellulaire pour laquelle de la benzonase est utilisée avec 10mM MgCl_2 . Lors du passage sur la colonne de nickel, le complexe Mlh1*Pms1 passe dans l'éluat direct et peut emporter avec lui des traces d'ions Ni^{2+} . Enfin à l'étape de cristallogenèse, seul 10mM MgCl_2 est utilisé pour améliorer les cristaux.

Afin d'identifier la nature de l'ion présent dans le site endonucléase, nous avons réalisé les spectres de fluorescence sur un cristal du CTD de MutL α . Ces spectres sont réalisés à des énergies proches des seuils d'absorption des métaux susceptibles de donner un signal anomal à la longueur d'onde utilisée pour le jeu SeMet. Nous avons ainsi réalisé des spectres de fluorescence autour des seuils d'absorption des atomes suivants : Ni, Fe, Cu, Co, As, Mn et Zn (**Figure 3 - 26**). Deux spectres présentent un pic d'absorption. Le spectre réalisé à une énergie proche de la transition K du zinc présente un fort signal d'absorption. Le spectre réalisé autour du seuil d'absorption du fer montre un signal cinq fois plus faible. Ce signal peut provenir de la tige en fer de la boucle supportant le cristal. Ces premiers spectres suggèrent que les cristaux contiennent principalement des cations Zn^{2+} .

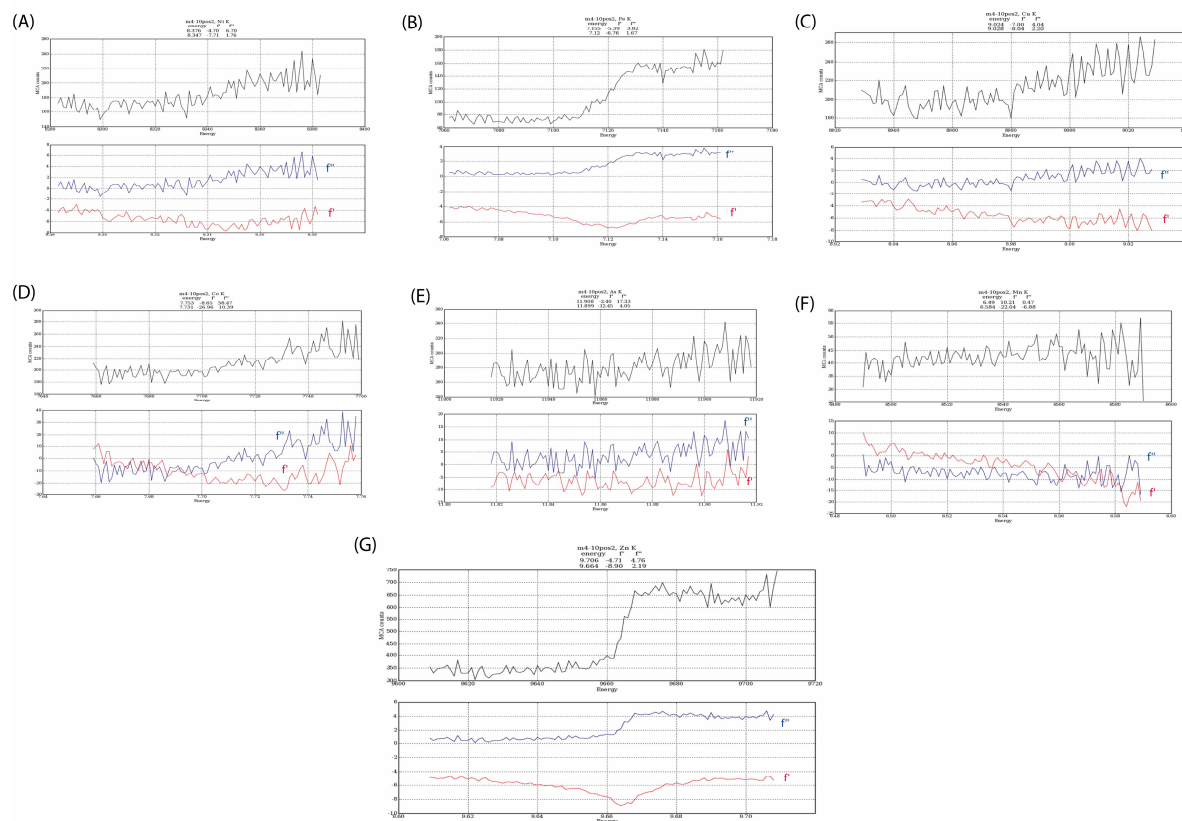


Figure 3 - 26: Spectres de fluorescence autour de l'énergie d'absorption de la transition K du (A) nickel (B) fer (C) cuivre (D) cobalt (E) arsenic (F) manganèse et (G) du zinc

Dans un second temps, nous avons collecté deux jeux de données sur un même cristal diffractant jusqu'à 4.5\AA (**Tableau 3 - 18**). Le premier jeu est réalisé au pic d'absorption de la transition K du zinc ($\lambda=1.2815\text{\AA}$) (**Figure 3 - 27(A)**). Le second jeu est collecté à une énergie précédant l'absorption ($\lambda=1.2861\text{\AA}$) (**Figure 3 - 27(B)**). Nous avons affiné la structure du complexe avec ces deux jeux de données. Le jeu effectué au pic d'absorption présente un fort pic anomal d'intensité supérieure à 11σ et similaire à ce que l'on observe dans le jeu réalisé avec le cristal SeMet. Aucun signal anomal n'est observé lorsque l'on se place juste en dessous du seuil d'absorption au zinc.

Tableau 3 - 18 : Statistiques des données de diffraction au seuil et avant le seuil de la transition K du zinc; les valeurs entre parenthèses correspondent aux statistiques dans la dernière coquille de résolution

	SeMet Pic Zn	SeMet Pré-pic Zn
Collection des données		
Groupe d'espace	C2	C2
Paramètres de maille		
a, b, c (Å)	190.9, 66.2, 74.7	190.9, 66.2, 74.7
α, β, γ (°)	90.0, 90.8, 90.0	90.0, 90.8, 90.0
Longueur d'onde (Å)	1.2815	1.2861
Résolution (Å)	35.0-4.55	35.0-4.53
	(4.97-4.55)	(4.88-4.53)
R_{sym} ou R_{merge}	18.7 (24.9)	30.2 (51.9)
$I / \sigma I$	6.10 (4.03)	4.06 (1.97)
Complétude (%)	98.2 (98.9)	98.4 (98.9)
Redondance	2.0 (2.0)	2.0 (2.0)

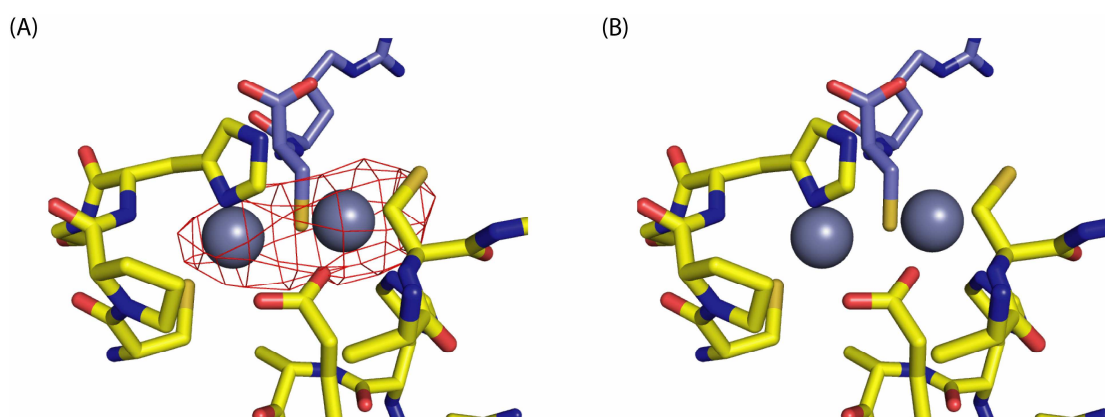


Figure 3 - 27 : Identification des ions Zn^{2+} (A) Représentation du signal anormal (grillage rouge) au seuil d'absorption du zinc ($\lambda=1.2815\text{\AA}$) ; les ions Zn^{2+} sont représentés par des sphères grises; le signal anormal dans la carte de Fourier-différence anormale est représenté à un niveau de 4σ pour une résolution comprise entre 30 et 4\AA (B) Représentation du signal anormal (grillage rouge) avant le seuil d'absorption du zinc ($\lambda=1.2861\text{\AA}$) ; les ions Zn^{2+} sont représentés par des sphères grises; aucun signal anormal n'est observé dans la carte de Fourier-différence anormale contourée à un niveau de 4σ pour une résolution comprise entre 30 et 4\AA

Sur la base de ces résultats, la structure native du complexe Mlh1-Pms1 a été affinée avec deux atomes de Zn^{2+} dans le site endonucléase de Pms1. Ces atomes sont affinés à 2.6\AA avec des facteurs d'agitation thermique, facteur B, de 45 et 48 contre 60\AA^2 en moyenne pour les acides aminés environnant de Pms1. Les pics de densité positive observés dans le site endonucléase de Pms1 dans la carte de densité électronique $F_o - F_c$ disparaissent avec l'introduction des deux atomes de Zn^{2+} .

Ces résultats confirment donc que ce site, prédit comme important pour l'activité endonucléasique, contient dans le cristal deux cations Zn^{2+} . Le site endonucléase possède vraisemblablement une forte affinité pour le Zn^{2+} car ces ions n'ont été introduits à aucune étape de la purification et de la cristallogenèse. De plus ces ions sont restés au niveau de ce site malgré la présence de 10mM MgCl_2 dans la cristallogenèse.

Structure du site métallique lié à l'activité endonucléase de Pms1

Les deux cations Zn^{2+} identifiés sont appelés Zn_1^{2+} et Zn_2^{2+} (**Figure 3 - 28**). Le premier atome, Zn_1^{2+} , est coordonné par 3 atomes de Pms1: l'oxygène $\text{O}_{\epsilon 1}$ du glutamate E707^{P} , le thiol S_{γ} de la cystéine C848^{P} et le $\text{N}_{\delta 1}$ de l'histidine H850^{P} . Le second atome, Zn_2^{2+} , est coordonné par 3 atomes de Pms1: l'oxygène $\text{O}_{\epsilon 2}$ du glutamate E707^{P} , le thiol S_{γ} de la cystéine C817^{P} et le $\text{N}_{\epsilon 2}$ de l'histidine H703^{P} . Les distances moyennes entre ces atomes sont standard pour des liaisons de sites de fixation du Zn (moyenne des distances = $2.25\text{\AA} \pm 0.25\text{\AA}$). On peut noter la symétrie du site de fixation des deux ions Zn^{2+} avec un thiol d'une cystéine, un groupement imidazole d'histidine pour chaque Zn^{2+} et le groupement carboxyle d'un glutamate qui pontre entre les deux atomes Zn^{2+} . La distance entre les deux ions zinc(II) est de 3\AA , ce qui est standard (de l'ordre de $3\text{-}4\text{\AA}$ pour un site catalytique avec deux zincs pontés par un glutamate ou un aspartate (Auld, 2001)).

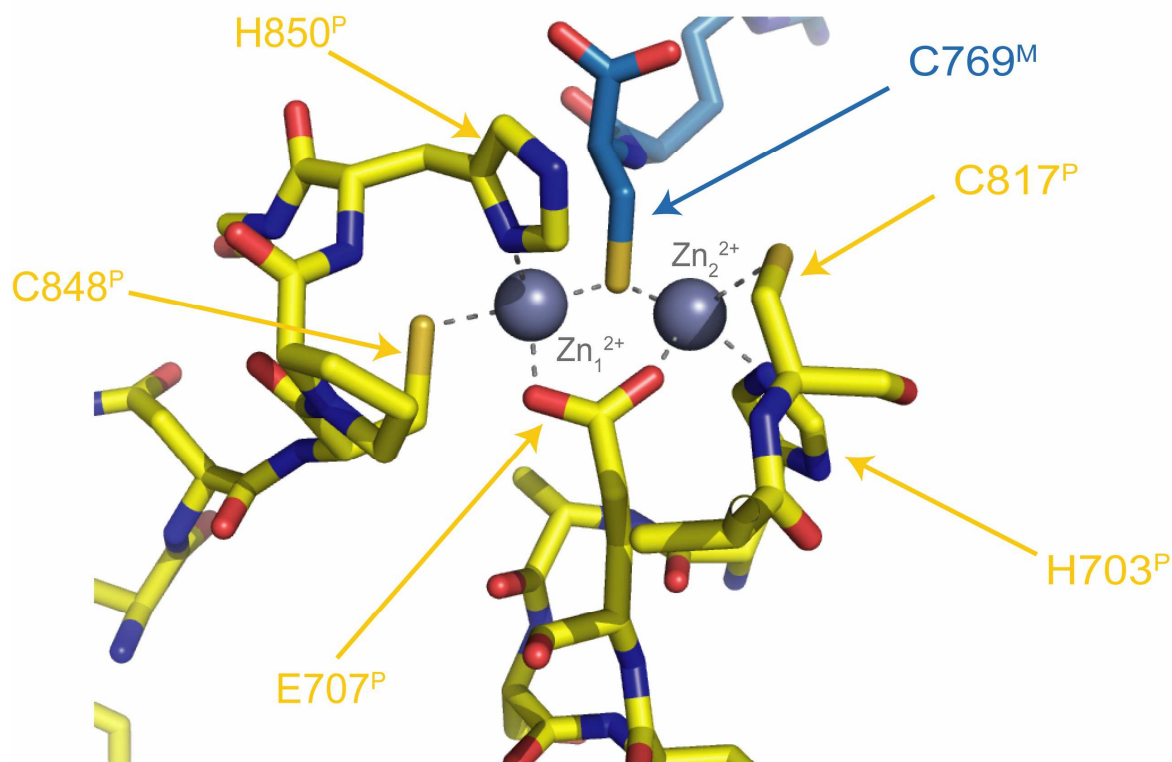


Figure 3 - 28: Structure du site endonucléasique du domaine C-terminal de Mlh1*Pms1; le CTD de Pms1 est représenté en jaune, le CTD de Mlh1 en bleu et les deux cations Zn^{2+} par des sphères grises

De façon surprenante, l'analyse de la structure du site endonucléase de Pms1 révèle la participation d'atomes de Mlh1 comme quatrième élément de coordination des deux atomes de Zn^{2+} . Comme cela avait été évoqué lors de la description du site d'hétérodimérisation, l'extrémité C-terminale de Mlh1 embrasse le domaine de dimérisation de Pms1 jusqu'à atteindre le site endonucléase. La structure révèle que

le dernier acide aminé de Mlh1, la cystéine C769, vient se positionner au-dessus des deux atomes de Zn^{2+} . (**Figure 3 - 29(A)**). L'analyse des cartes de densité électronique suggère deux scénarios possibles quant à la nature des atomes de la cystéine C769 qui chélatent les deux atomes de zinc. A la résolution de 2.6Å, les cartes sont compatibles avec le positionnement du thiol de C769 vers les deux atomes de Zn^{2+} comme représenté sur la figure ci-dessous (**Figure 3 - 28**). Les cartes sont également compatibles avec le positionnement de la fonction carboxyle terminale vers les deux atomes de Zn^{2+} (**Figure 3 - 29(B)&(C)**). Le thiol de la cystéine et l'extrémité COOH partagent un même nombre d'électrons (17 pour le thiol SH et 17 pour le groupement COO⁻ sous sa forme chargée). Ceci rend difficile la discrimination des deux scénarios d'après les seules cartes de densité électronique $2mF_o - F_c$ (**Figure 3 - 29(B)&(C)**).

Mutagenèse du site endonucléase de Pms1 et de l'extrémité C-terminale de Mlh1: A cette étape, nous pouvons re-analyser les données de la littérature par rapport aux informations structurales obtenues sur le site endonucléase de Pms1. La structure cristallographique confirme les prédictions issues des analyses de séquences puis des premières structures cristallographiques de régions C-terminales de MutL bactériens à activité endonucléase. Les trois motifs proposés sont bien impliqués dans la fixation des ions Zn^{2+} métalliques observés dans le cristal. Notamment la perte d'activité nucléase de Pms1 observée avec le mutant E707K^P (Kadyrov et al., 2007) est en bon accord avec le rôle central de la position E707 dans la coordination des deux atomes de zinc. Du côté de Mlh1, la structure cristallographique permet de mieux comprendre l'importance fonctionnelle de l'extrémité C-terminale de Mlh1. La construction Mlh1ΔC10, raccourcie des dix derniers acides aminés interagit encore en test de double-hybride avec Pms1 comme évoqué dans le paragraphe sur le site d'hétérodimérisation mais complémente une souche *mlh1Δ* beaucoup plus faiblement qu'un plasmide contenant Mlh1 sauvage (expérience faite avec un essai qui mesure le taux de l'allèle *hom3-10*) (Pang et al., 1997). La position de l'extrémité C-terminale de Mlh1 à proximité du site endonucléase suggère que cette délétion perturbe significativement le site endonucléase de Pms1 ou la reconnaissance de l'ADN à cliver.

Dans la même étude, les auteurs ont analysé différentes mutations ponctuelles de l'extrémité C-terminale de Mlh1 introduites sur le chromosome de la levure. Tous les mutants ponctuels testés présentent une interaction en double hybride avec Pms1 similaire à la protéine sauvage. Trois mutants de cette zone présentent un effet mutateur très fort. Il s'agit des mutants K764E^M, F766A^M et E767D^M qui présentent des taux de mutation relatifs par rapport aux souches sauvages d'un facteur 600 à 970 comparé au facteur 1160 observé pour la souche *mlh1Δ* (essai par mesure du taux de reversion de l'allèle *hom3-10*). L'acide aminé K764^M effectue une liaison intramoléculaire avec D761^M. L'acide aminé F766^M est enfoui à l'interface avec Pms1. L'acide aminé E767^M fait un pont salin intermoléculaire avec R852^P de Pms1. La structure suggère que les trois mutations analysées par Pang et al. affectent l'hélice à

l'extrémité C-terminale de Mlh1 (K764^M) ou bien l'interaction de cette hélice avec Pms1 (F766A^M et E767D^M). Ces perturbations ne semblent pas affecter l'interaction globale avec Pms1 mais sont susceptibles de perturber la géométrie du site endonucléase. De façon intéressante deux mutants ponctuels ont été testés par Pang et al. sur la dernière position de Mlh1.

Analyse des séquences des extrémités C-terminales des protéines Mlh1: Les alignements multiples des séquences eucaryotes de Mlh1 montrent une conservation très forte des dix derniers acides aminés de Mlh1 pour la grande majorité des séquences³⁵. Il est à noter que chez *C. elegans* et *C. briggsae*, la séquence de Mlh1 est plus longue de deux acides aminés (FERCGT₇₅₈ chez *C. elegans* et FERCGS₇₆₅ chez *C. briggsae*, le motif souligné représentant les 4 derniers acides aminés de la plupart des séquences Mlh1). Dans l'hypothèse où *C. elegans* et *C. briggsae* ne représentent pas un cas unique parmi tous les eucaryotes, cette observation suggère fortement que la chaîne latérale de la cystéine des Mlh1 est le groupement qui chélate les atomes de zinc. En effet, la protéine Mlh1 de *C. elegans* et *C. briggsae* ne présente plus de fonction carboxyle en position 769 chez la levure. De façon intéressante, deux mutants ponctuels ont été testés par Pang et al. sur la dernière position de Mlh1. Il s'agit des mutants C769A^M et C769S^M qui présentent respectivement aucun phénotype mutateur ou un phénotype mutateur fort (taux de mutation de 290). Dans le modèle que nous proposons aujourd'hui d'une participation du thiol de la cystéine C769 dans la chélation des zinc, l'absence ce phénotype mutateur pour le mutant C769A^M pourrait s'expliquer par un changement de conformation permis par la petite taille de l'alanine qui pourrait occuper la place du groupement carboxyle. Ce dernier occuperait la position de la cystéine pour chelater les deux atomes de zinc. Dans notre modèle, le phénotype mutateur observé avec le mutant C769S^M correspondrait à une chélation imparfaite par la fonction hydroxyle de la sérine. On peut noter dans ce sens, l'absence de cet acide aminé dans la grande majorité des sites de fixation du zinc.

³⁵ L'alignement de séquence du domaine C-terminal de Mlh1 est présenté en ANNEXE 16

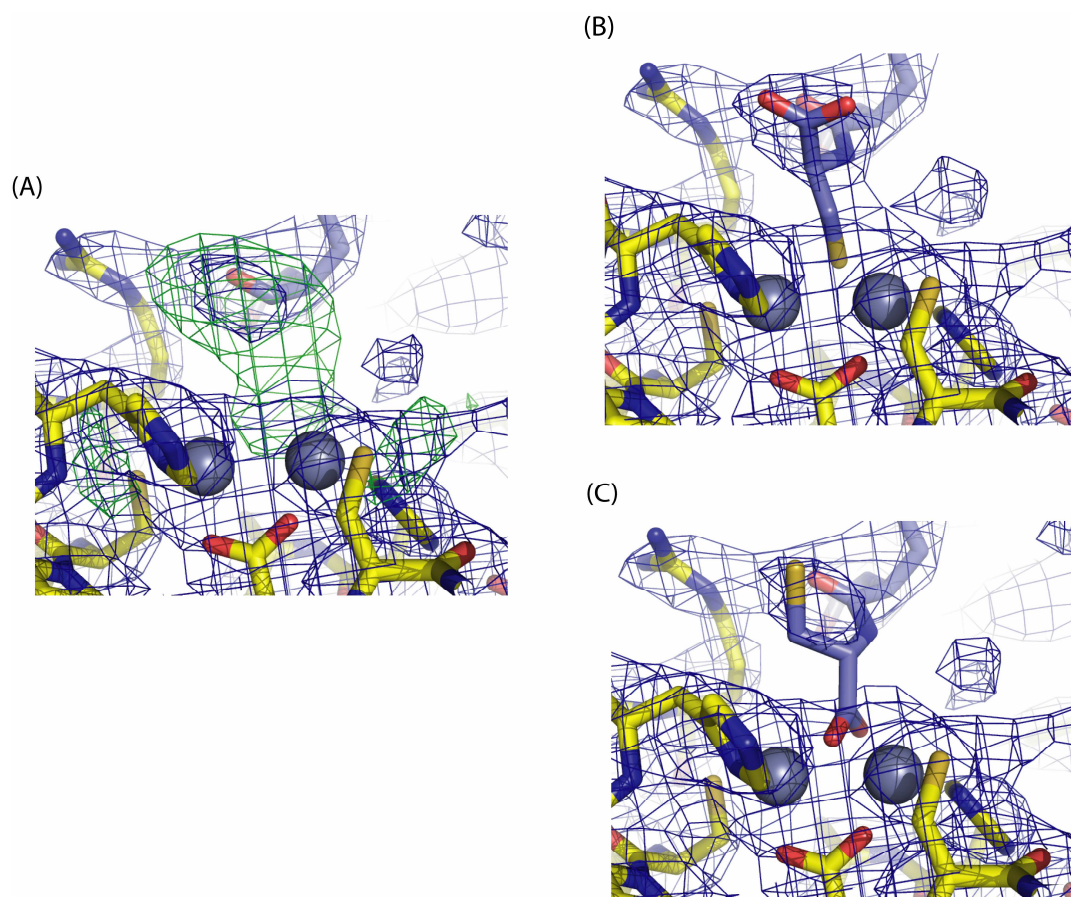


Figure 3 - 29 : Rôle du dernier acide aminé de Mlh1, C769, dans le site endonucléase de Pms1 (A) Carte *omit* de densité électronique du résidu C769 de Mlh1 ; la carte $2mF_o-F_c$ est contournée à 1.3σ (bleu), la carte F_o-F_c (vert) est contournée à $+3\sigma$ (B) Carte de densité électronique autour du résidu C769 de Mlh1 avec l'atome de soufre impliqué dans le site métallique; la carte $2mF_o-F_c$ est contournée à 1.3σ (bleu) (C) Carte de densité électronique du résidu C769 de Mlh1 avec l'extrémité COOH impliqué dans le site métallique; la carte $2mF_o-F_c$ est contournée à 1.3σ (bleu)

En conclusion, la structure du site endonucléase de Pms1 confirme les prédictions publiées précédemment quant aux rôles des trois motifs conservés. Le site endonucléase est occupé dans le cristal par deux atomes de zinc chélatés par cinq acides aminés de Pms1 (H703, E707, C817, C848, H850) et le dernier acide aminé de Mlh1 (C769). La structure de ce site est en bon accord avec les données de mutagenèse de la littérature sur les régions de Pms1 impliquées dans ce site. La structure du complexe Mlh1(CTD)*Pms1(CTD) représente la première mise en évidence de l'interaction directe de Mlh1 sur la structure du site endonucléase de Pms1.

4. CONCLUSIONS ET PERSPECTIVES

Au cours de la première partie de ce travail de thèse, l'objectif, que nous nous étions fixé, était de caractériser le site d'interaction de la protéine Mlh1 de levure avec une famille de protéines partageant un même motif d'interaction avec Mlh1 appelé motif MIP pour Mlh1 interacting protein motif. Ce motif, mis en évidence dans le laboratoire de S. Boiteux, est constitué des cinq acides aminés (R/K)SK(Y/F)F. Le motif MIP est présent sur les protéines Exo1, Ntg2 et Sgs1 de levure et EXO1 et BLM humaines qui interviennent dans des voies de réparation différentes. Notre objectif était de caractériser le ou les sites d'interaction de Mlh1 avec les protéines contenant un motif MIP et de localiser ce site par rapport au site d'hétérodimérisation de Mlh1 avec les protéines Pms1, Mlh2 et Mlh3.

Nous avons pu proposer par une approche combinant mutagenèse dirigée, double hybride et modélisation moléculaire, un ensemble de dix acides aminés susceptibles de constituer le site d'interaction des protéines Ntg2 et Exo1 avec Mlh1, appelé site S2. Les acides aminés de ce site S2 ne sont pas impliqués dans l'interaction avec Pms1 qui met en jeu un autre site, appelé site S1. Ces résultats ont été confirmés par des mesures de calorimétrie entre la région C-terminale de levure et des peptides dérivés des protéines Ntg2, Exo1 et Sgs1. Nous avons pu également montrer que cette interaction est conservée entre la région C-terminale de MLH1 humain et des peptides dérivés des protéines EXO1 et BLM humaines. Afin d'évaluer l'importance du site S2 dans la voie de réparation MMR, un mutant chromosomique de ce site a été caractérisé. Il a pu être observé que ce mutant *mlh1-E682A* présente un phénotype mutateur proche d'une souche *exo1Δ* que ce soit dans un contexte sauvage ou dans un contexte d'un mutant hypomorphe du MMR (*pms1-G97A*). Plus récemment, un rôle du site S2 de Mlh1 dans les mécanismes de recombinaison méiotique a été mis en évidence dans le laboratoire de N. Hunter (Zakharyevich et al., 2010). Ces auteurs ont montré que le mutant *mlh1-E682A* du site S2 entraîne une diminution du nombre de crossover et que cette diminution est due à un défaut de recrutement d'Exo1 par son motif MIP.

Dans un second temps, nous avons souhaité caractériser par cristallographie, la structure de la région C-terminale de Mlh1 de levure. La stratégie qui s'est avérée la plus efficace a été la cristallisation de la région C-terminale de l'hétérodimère Mlh1*Pms1 (MutL α). La structure cristallographique a été résolue par marquage de la protéine avec des sélénométhionines. La région C-terminale de MutL α présente un repliement général semblable à celui de la région C-terminale des protéines MutL bactériennes. La surface d'hétérodimérisation implique un empilement d'un feuillet β de Mlh1 sur un feuillet β de Pms1. Cette interface rappelle l'interface d'homodimérisation des MutL bactériens. En revanche, la surface

d'hétérodimérisation implique également les extrémités C-terminales de Mlh1 et Pms1 selon un mode d'interaction différent des protéines bactériennes. La structure de Mlh1 valide les prédictions faites au cours de la première étude sur l'existence d'un site S2 et sur la nature des acides aminés qui le constituent. La forme du site S2 permet de proposer un mode d'interaction des peptides contenant un motif MIP sur ce site. Enfin l'analyse du site endonucléase de Pms1 a permis de préciser le rôle des motifs du site endonucléase identifiés par les analyses bioinformatiques. De façon inattendue, la structure cristallographique de Mlh1*Pms1 indique la présence de deux atomes de zinc au sein du site endonucléase et le rôle du dernier acide aminé de Mlh1, C769 dans la coordination du zinc.

Dans la dernière partie de ce manuscrit, nous avons choisi de présenter différents travaux qui nous paraissent intéressants d'explorer dans le futur soit parce qu'ils sont un prolongement direct des travaux présentés dans la partie précédente, soit parce qu'ils correspondent à des travaux pour lesquels j'ai initié les études sans pouvoir les mener à leur terme ou enfin parce qu'ils correspondent à des questions nouvelles importantes.

Comment MutL α encercle-t-elle l'ADN en vue de son incision?

La présence d'une activité endonucléase à la surface du domaine C-terminal de MutL α implique la fixation de l'ADN à proximité de ce site actif. Pour guider l'identification de ce site, nous avons calculé le potentiel électrostatique à la surface de la région C-terminale de Mlh1*Pms1 à l'aide du serveur PBEQ (**Figure 4 - 1(A)**). L'analyse de ce potentiel montre qu'une des faces de Pms1(CTD) est fortement chargée positivement. Cette surface basique s'étend aussi bien sur le domaine de dimérisation au niveau de l'hélice α_E que sur tout le flan du domaine régulateur. Afin d'avoir une première idée du mode d'interaction de l'ADN avec la région C-terminale de MutL α , nous avons réalisé un premier ancrage manuel d'un double brin d'ADN dans cette région du complexe (**Figure 4 - 1(B)**). Cet ancrage manuel positionne le site endonucléase à moins de 10Å d'une liaison phosphodiester d'ADN.

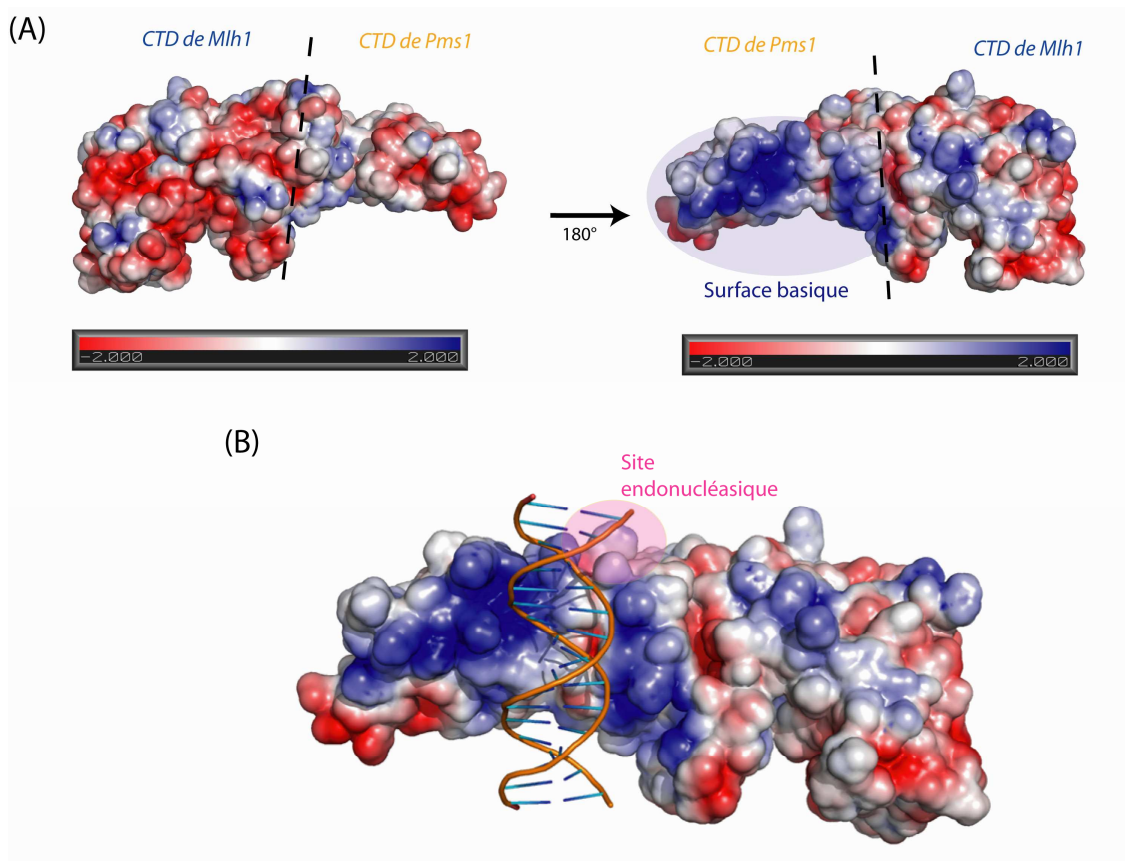


Figure 4 - 1 : Analyse électrostatique du domaine C-terminal de MutL α (A) Représentation de la surface du CTD de Mlh1*Pms1 colorée selon le potentiel électrostatique calculé par le serveur PBEQ; les résidus chargés négativement sont représentés en rouge, les résidus chargés positivement sont représentés en bleu (B) Ancrage manuel d'un brin d'ADN au sein de la surface convexe chargée positivement

L'analyse quantitative ou structurale de l'interaction de MutL α avec l'ADN semble délicate avec la seule région C-terminale. Nos collaborateurs du CEA de Fontenay-aux-Roses ont analysé l'activité endonucléase ou les propriétés de fixation de l'ADN de la région C-terminale de MutL α utilisée en cristallogénèse. Ces premières analyses ont mis en évidence une très faible activité endonucléase pour la région C-terminale seule et des propriétés de fixation d'un ADN double brin de 35 paires de bases faible. Ces données sont en accord avec les données de la littérature. Les MutL bactériens à activité endonucléase sont actifs sous leur forme entière alors que leur région C-terminale est peu active.

Aujourd'hui, afin d'étudier plus en avant l'étape d'incision par la protéine MutL α de l'ADN, il semble important de purifier la protéine MutL α entière dans *E. coli* ou en cellules d'insectes. L'obtention de cristaux de la protéine entière représenterait une étape majeure dans la compréhension de sa fonction. Les études de microscopie à force atomique suggèrent qu'une étape importante pour la cristallisation d'un complexe MutL α *ADN consiste à obtenir une forme stable de la protéine en jouant notamment sur les analogues de nucléotides présents (Sacho et al., 2008). On peut également penser à l'utilisation de DARPin (pour Designed Ankyrin Repeat Proteins). Cette approche développée par A. Plueckthun (University of

Zurich) et accessible par le programme européen PCUBE. Cette approche permet de sélectionner des protéines avec des répétitions de domaine ankryrin pour stabiliser une conformation donnée d'une protéine modulaire ou dynamique. Le complexe MutL α *ADN représente un exemple intéressant de cible pour cette approche.

S'il existe, où se cache donc le motif PIP (PCNA Interacting Protein motif) de MutL α ?

Le trimère de PcnA a été trouvé en 1996 dans un crible en double hybride de banque de levure contre la protéine Mlh1 comme appât (Umar et al., 1996). Dans cette même étude les auteurs ont mis en évidence un phénotype mutateur sur une séquence de 51 paires de bases de GT répétés pour un mutant de PcnA de levure, *pol30-52*, proche du taux de mutation observé pour des souches *mlh1 Δ* et du double mutant *pol30-52 mlh1 Δ* , suggérant que ces mutants reflètent un défaut du même processus. Les auteurs ont montré sur les protéines humaines, l'apparition d'un phénotype mutateur en présence d'un peptide p21 qui correspond à un motif PIP très affin de PCNA humain. Il a été proposé que l'interaction PCNA*MLH1 implique un motif PIP de MLH1. A partir de l'analyse de la séquence de Mlh1 de levure et de la recherche sur cette séquence des motifs PIP potentiels, Lee et al. ont récemment proposé que le motif d'interaction de Mlh1 avec PcnA était le motif ⁵⁷²QIGLTDEA. Ce motif possède trois des quatre positions importantes du motif canonique Qxxhxxaa où h est un résidu hydrophobe et a un résidu aromatique. Plusieurs éléments nous suggèrent aujourd'hui que ce motif n'est pas le motif fonctionnel et que le motif fonctionnel reste à identifier. (i) Premièrement, l'analyse de notre structure cristallographique montre que ce motif de Mlh1 de levure se situe à la fin de l'hélice α_A et que la phénylalanine du motif est complètement enfouie. La structure de la région ⁵⁷²QIGLTDEA de Mlh1 diverge fortement de la structure canonique des motifs PIP, caractérisés par un brin β comprenant la glutamine, suivi d'une hélice comprenant le résidu hydrophobe et les deux résidus aromatiques. (ii) Deuxièmement, nous avons testé au laboratoire par calorimétrie l'interaction entre la région C-terminale de Mlh1 de levure purifiée et un peptide contenant le motif PIP centré autour de Q572 proposé par Lee et al.. Nous n'avons observé aucune interaction avec du PcnA de levure. Par la même approche au laboratoire, nous avons pu caractériser l'interaction de plusieurs peptides contenant un PIP dont notamment celui d'EXO1 humain (motif ⁷⁸⁸QIKLNE \underline{L} W) (Liberti et al., 2010). (iii) Troisièmement, dans leur étude, Lee et al. n'observent un affaiblissement de l'interaction que pour le triple mutant Mlh1(QLF \rightarrow AAA). Un simple mutant de la glutamine est normalement suffisant pour affaiblir fortement l'interaction dans les autres interactions PCNA*PIP. Sur la base de ces observations, nous avons re-initié l'étude de l'interaction des différents PIP potentiels de Mlh1 afin d'évaluer la présence ou non d'un motif PIP canonique sur la séquence de Mlh1. Nous avons décidé d'élargir la recherche d'un motif PIP à la région N-terminale et à la région charnière entre les deux régions N- et

C-terminales de Mlh1 mais également à la protéine Pms1. L'existence de seulement deux articles sur l'interaction Mlh1*Pcna dont une analyse par double hybride et une analyse par résonance plasmonique de surface sur les triples mutants, suggère également que d'autres expériences (pull-down, co-immunoprécipitation) seraient nécessaires pour confirmer définitivement la réalité de l'interaction Mlh1*PCNA.

Existe-t-il d'autres protéines possédant un motif MIP chez la levure et chez l'homme?

Les travaux présentés dans cette thèse autour du site 2 de Mlh1 ont permis de mettre en évidence le rôle prépondérant au sein du motif MIP, **(R/K)SK(Y/F)F** de la sérine et des deux aromatiques. Parmi les nombreux partenaires de Mlh1 de levure ou humain, existent-il d'autres partenaires de Mlh1 susceptibles d'interagir *via* un motif MIP? Le tableau ci-dessous présente quelques partenaires pour lesquels nous avons commencé cette analyse (**Tableau 4 - 1**). Nous avons alors recherché dans leur séquence, la présence d'un ou plusieurs motifs MIP. Les trois premiers MIP potentiels testés ont été deux motifs potentiels de l'hélicase FancJ humaine et un motif de l'angiomotine humaine. L'interaction de FancJ avec MLH1 humain a été montrée par co-IP (Peng et al., 2007). L'interaction de l'angiomotine a été identifiée à partir d'un crible en Tap-Tag et confirmée par immunoprécipitation (Cannavo et al., 2007). Les mesures par calorimétrie des interactions de la région C-terminale de MLH1 humain purifiée avec des peptides contenant les séquences ₁₀₁NQGT**TSRHF**NYPS ou ₁₁₂₇DE**SIYF**TPEL de FancJ et la séquence ₁₁₂AKV**QSQYF**RGQQ de l'angiomotine ont montré l'absence d'interaction de ces peptides avec MLH1. Ce résultat négatif suggère que ces protéines interagissent par un autre mode d'action. De façon intéressante, ces résultats négatifs apportent des informations sur la spécificité des interactions de Mlh1 avec le motif MIP. Il semble en effet que la présence d'un acide aspartique et d'une isoleucine de part et d'autre de la sérine comme dans l'angiomotine ou d'une glutamine de part et d'autre de la sérine comme dans FancJ gène fortement l'interaction. Quatre autres peptides dérivés de la séquence de polη et Rad9 humain sont en cours d'analyse au laboratoire. L'interaction de la protéine polη a été montrée par co-IP (Kanao et al., 2009) et celle de Rad9 par co-IP (He et al., 2008).

Tableau 4 - 1: Séquence des peptides de partenaires de Mlh1 de levure ou humain et contenant un motif ressemblant à un motif MIP

Partenaire	Séquence du peptide	Interaction par calorimétrie
scMgs1 n°1	²⁰² NVGSRYFMIETSATK	Pas d'interaction avec le CTD de Mlh1
scMgs1 n°2	¹⁶ ICRSRKVFVFSL	à tester
hsFancJ n°1	¹⁰¹ NQGTSRHFENYPS	Pas d'interaction avec le CTD de MLH1
hsFancJ n°2	¹¹²⁷ DESIYFTPEL	Pas d'interaction avec le CTD de MLH1
Angiomotine	¹¹² AKVQSQYFRGQQ	Pas d'interaction avec le CTD de MLH1
hsMED1(MBD4)	⁴¹¹ RRKTSLYFSSKY	à tester
hsRAD9 n°1	²³⁵ NLSIHFDAP	à tester
hsRAD9 n°2	³⁶¹ FRSLFFGSIL	à tester
hsPol η n°1	¹⁰¹ EIMSRFAVIE	à tester
hsPol η n°2	³²⁰ GCSKNFPG	à tester
hsPol η n°3	³⁹⁴ KMSHDAFTV	à tester
hsPol η n°4	⁴⁸⁰ LESFFQKA	à tester

Comment Pms1 et Mlh3 se partagent-ils Mlh1 pour former MutL α et MutL γ ?

La structure cristallographique de la région C-terminale de Mlh1*Pms1 suggère que l'hétérodimère Mlh1*Mlh3 présente de grande similarité sur la surface d'hétérodimérisation notamment en ce qui concerne l'interaction entre les deux feuillets β et l'interaction de l'extrémité C-terminale de Mlh1. En revanche, les interactions observées avec l'extrémité C-terminale de Pms1 sont, selon toute vraisemblance, assez différentes avec Mlh3. Les alignements de séquences des Pms1 eucaryotes et des Mlh3 eucaryotes (ANNEXE 17) montrent que la séquence de Mlh3 est plus courte que celle de la protéine Pms1; ainsi, la séquence de Mlh3 s'arrête 11 acides aminés avant la fin de la séquence de Pms1. Ceci signifie que de nombreuses interactions comme le pont salin entre R547^M et D870^P ne sont pas présentes dans le complexe Mlh1*Mlh3. Il serait important de résoudre la structure cristallographique de la région C-terminale d'un complexe Mlh1*Mlh3 de levure ou humain afin d'analyser l'impact de cette différence importante au niveau du site d'hétérodimérisation sur la structure globale du complexe. Les données structurales maintenant disponibles sur le complexe C-terminal de Mlh1*Pms1 devrait permettre de définir des constructions plus précises des protéines Mlh1 et Mlh3 pour la cristallogénèse. La structure du complexe Mlh1*Mlh3 permettrait de mieux comprendre la spécificité de l'interaction de Mlh1 pour ces deux partenaires. Elle permettrait également d'analyser la position du domaine régulateur avec une autre protéine et dans un autre contexte cristallin.

Peut-on utiliser les informations structurales et fonctionnelles sur Mlh1 de levure pour une meilleure compréhension de l'impact des mutations ponctuelles associées au syndrome de Lynch?

Un enjeu important en diagnostic clinique est une meilleure compréhension de l'impact des mutations ponctuelles observées sur les protéines associées à des cancers. Notre laboratoire est impliqué dans deux projets autour de cette thématique. Il s'agit de l'analyse des mutations identifiées sur les facteurs de réparation des cassures doubles-brins, Cernunnos et Artemis, en collaboration avec le laboratoire de JP de Villartay à l'Hôpital Necker. Des patients ont ainsi été identifiés portant des mutations ponctuelles sur ces protéines et présentant des déficits majeurs du système immunitaire qui utilise de façon programmée la réparation des cassures doubles-brins (de Villartay et al., 2009; Malivert et al., 2010). Plus récemment, en collaboration avec les laboratoires de R. Lidereau (Centre René Huguenin (CRH), Institut Curie, St-Cloud) un projet a été initié par S. Zinn-Justin au laboratoire pour analyser les mutations ponctuelles de la protéine BRCA1, protéine associée aux cancers du sein. L'objectif est d'analyser par des méthodes biophysiques, un nombre important de variants de BRCA1, recensés sur la base de données gérée par le CRH, pour lesquels l'impact de la mutation sur la fonction de BRCA1 n'est pas connu. Il me paraît très important d'initier avec le CRH qui travaille également sur Mlh1, une approche similaire en tenant compte des informations structurales sur les régions N- et C-terminales de Mlh1. La forte similarité de séquences entre les régions C-terminales de levure et humaines (similarité de 53%) permet de construire un modèle par homologie fiable et de localiser les mutations ponctuelles associées à un phénotype sévère, intermédiaire ou faible. Il sera notamment intéressant d'étudier plus en avant par des méthodes biophysiques (calorimétrie, dichroïsme circulaire voire cristallographie) les mutants correspondant à une protéine Mlh1 correctement repliée et localisée au noyau mais non fonctionnelle.

- Abdullah, M.F., Hoffmann, E.R., Cotton, V.E. and Borts, R.H. (2004) A role for the MutL homologue MLH2 in controlling heteroduplex formation and in regulating between two different crossover pathways in budding yeast. *Cytogenet Genome Res*, **107**, 180-190.
- Acharya, S., Foster, P.L., Brooks, P. and Fishel, R. (2003) The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. *Mol Cell*, **12**, 233-246.
- Afonine, P.V., Grosse-Kunstleve, R.W. and Adams, P.D. (2005) The Phenix refinement framework. *CCP4 Newsl.* **42**.
- Afonine, P.V., Grosse-Kunstleve, R.W., Chen, V.B., Headd, J.J., Moriarty, N.W., Richardson, J.S., Richardson, D.C., Urzhumtsev, A., Zwart, P.H. and Adams, P.D. (2010) phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *Journal of Applied Crystallography*, **43**, 669-676.
- Ahrends, R., Kosinski, J., Kirsch, D., Manelyte, L., Giron-Monzon, L., Hummerich, L., Schulz, O., Spengler, B. and Friedhoff, P. (2006) Identifying an interaction site between MutH and the C-terminal domain of MutL by crosslinking, affinity purification, chemical coding and mass spectrometry. *Nucleic Acids Res*, **34**, 3169-3180.
- Alam, N.A., Gorman, P., Jaeger, E.E.M., Kellsell, D., Leigh, I.M., Ratnavel, R., Murdoch, M.E., Houlston, R.S., Aaltonen, L.A., Roylance, R.R. and Tomlinson, I.P.M. (2003) Germline deletions of EXO1 do not cause colorectal tumors and lesions which are null for EXO1 do not have microsatellite instability. *Cancer Genetics and Cytogenetics*, **147**, 121-127.
- Alani, E., Sokolsky, T., Studamire, B., Miret, J.J. and Lahue, R.S. (1997) Genetic and biochemical analysis of Msh2p-Msh6p: role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition. *Mol Cell Biol*, **17**, 2436-2447.
- Ali, J.A. and Lohman, T.M. (1997) Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase. *Science*, **275**, 377-380.
- Allen, D.J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P. and Griffith, J.D. (1997) MutS mediates heteroduplex loop formation by a translocation mechanism. *Embo J*, **16**, 4467-4476.
- Aloy, P. and Russell, R.B. (2006) Structural systems biology: modelling protein interactions. *Nat Rev Mol Cell Biol*, **7**, 188-197.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
- Amin, N.S., Nguyen, M.N., Oh, S. and Kolodner, R.D. (2001) *exo1*-Dependent mutator mutations: model system for studying functional interactions in mismatch repair. *Mol Cell Biol*, **21**, 5142-5155.
- An, L., Tang, W., Ranalli, T.A., Kim, H.J., Wytiaz, J. and Kong, H. (2005) Characterization of a thermostable UvrD helicase and its participation in helicase-dependent amplification. *J Biol Chem*, **280**, 28952-28958.

- Andrade, M.A., Chacon, P., Merelo, J.J. and Moran, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng*, **6**, 383-390.
- Antony, E. and Hingorani, M.M. (2004) Asymmetric ATP binding and hydrolysis activity of the *Thermus aquaticus* MutS dimer is key to modulation of its interactions with mismatched DNA. *Biochemistry*, **43**, 13115-13128.
- Arana, M.E., Holmes, S.F., Fortune, J.M., Moon, A.F., Pedersen, L.C. and Kunkel, T.A. (2010) Functional residues on the surface of the N-terminal domain of yeast Pms1. *DNA Repair (Amst)*, **9**, 448-457.
- Argueso, J.L., Kijas, A.W., Sarin, S., Heck, J., Waase, M. and Alani, E. (2003) Systematic Mutagenesis of the *Saccharomyces cerevisiae* MLH1 Gene Reveals Distinct Roles for Mlh1p in Meiotic Crossing Over and in Vegetative and Meiotic Mismatch Repair. *Mol Cell Biol*, **23**, 873-886.
- Asherie, N. (2004) Protein crystallization and phase diagrams. *Methods*, **34**, 266-272.
- Auld, D.S. (2001) Zinc coordination sphere in biochemical zinc sites. *Biometals*, **14**, 271-313.
- Baker, S.M., Bronner, C.E., Zhang, L., Plug, A.W., Robatzek, M., Warren, G., Elliott, E.A., Yu, J., Ashley, T., Arnheim, N., Flavell, R.A. and Liskay, R.M. (1995) Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell*, **82**, 309-319.
- Ban, C., Junop, M. and Yang, W. (1999) Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*, **97**, 85-97.
- Ban, C. and Yang, W. (1998a) Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell*, **95**, 541-552.
- Ban, C. and Yang, W. (1998b) Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases. *Embo J*, **17**, 1526-1534.
- Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M. and Neri, G. (1999) MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc Natl Acad Sci U S A*, **96**, 3969-3974.
- Bjornson, K.P., Allen, D.J. and Modrich, P. (2000) Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry*, **39**, 3176-3183.
- Blundell, T.L. and Johnson, L.N. (1976) *Protein crystallography*. Academic Press (New York).
- Bonvin, A.M. (2006) Flexible protein-protein docking. *Curr Opin Struct Biol*, **16**, 194-200.
- Bornhop, D.J., Latham, J.C., Kussrow, A., Markov, D.A., Jones, R.D. and Sorensen, H.S. (2007) Free-solution, label-free molecular interactions studied by back-scattering interferometry. *Science*, **317**, 1732-1736.

- Bragg, W.L. (1913) The diffraction of short electromagnetic waves by a crystal. *Proc. Camb. Philos. Soc.*, **17**, 43-57.
- Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courcon, M., Masson, C., Dassa, J., Charbonnier, J.B., Boulain, J.C., Menez, A., Genet, R. and Gondry, M. (2005) Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J Proteome Res*, **4**, 2137-2147.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vornrhein, C. and Womack, T. (2009) *BUSTER version 2.9*, Global Phasing Ltd., Cambridge, United Kingdom.
- Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M. and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem*, **30**, 1545-1614.
- Brown, J., Brown, T. and Fox, K.R. (2001) Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. *Biochemical Journal*, **354**, 627-633.
- Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S.T. and Modrich, P. (2001) In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci U S A*, **98**, 6765-6770.
- Calin, G., Ranzani, G.N., Amadori, D., Herlea, V., Matei, I., Barbanti-Brodano, G. and Negrini, M. (2001) Somatic frameshift mutations in the Bloom syndrome BLM gene are frequent in sporadic gastric carcinomas with microsatellite mutator phenotype. *BMC Genet*, **2**, 14.
- Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R. and Jiricny, J. (2007) Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J Biol Chem*, **282**, 2976-2986.
- Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S.M., Fischer, F., Cejka, P. and Jiricny, J. (2005) Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res*, **65**, 10759-10766.
- Cao, A. (2003) Light scattering. Recent applications. *Analytical Letters*, **36**, 3185-3225.
- Chao, E.C., Velasquez, J.L., Witherspoon, M.S., Rozek, L.S., Peel, D., Ng, P., Gruber, S.B., Watson, P., Rennert, G., Anton-Culver, H., Lynch, H. and Lipkin, S.M. (2008) Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). *Hum Mutat*, **29**, 852-860.
- Charbonneau, N., Amunugama, R., Schmutte, C., Yoder, K. and Fishel, R. (2009) Evidence that hMLH3 functions primarily in meiosis and in hMSH2-hMSH3 mismatch repair. *Cancer Biol Ther*, **8**, 1411-1420.
- Chen, P.C., Dudley, S., Hagen, W., Dizon, D., Paxton, L., Reichow, D., Yoon, S.R., Yang, K., Arnheim, N., Liskay, R.N. and Lipkin, S.M. (2005) Contributions by MutL homologues MLH3

- and PMS2 to DNA mismatch repair and tumor suppression in the mouse. *Cancer Research*, **65**, 8662-8670.
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography*, **66**, 12-21.
- Cole, J.L. and Garsky, V.M. (2001) Thermodynamics of peptide inhibitor binding to HIV-1 gp41. *Biochemistry*, **40**, 5633-5641.
- Constantin, N., Dzantiev, L., Kadyrov, F.A. and Modrich, P. (2005) Human mismatch repair: reconstitution of a nick-directed bidirectional reaction. *J Biol Chem*, **280**, 39752-39761.
- Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Crystallographica Section D-Biological Crystallography*, **64**, 83-89.
- Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallographica Section D-Biological Crystallography*, **66**, 470-478.
- Crick, F. (1974) The double helix: a personal view. *Nature*, **248**, 766-769.
- Crippen, G.M. and Havel, T.F. (1990) Global energy minimization by rotational energy embedding. *J Chem Inf Comput Sci*, **30**, 222-227.
- Cromer, D.T. (1965) Anomalous Dispersion Corrections Computed from Self-Consistent Field Relativistic Dirac-Slater Wave Functions. *Acta Crystallographica*, **18**, 17-&.
- Cutalo, J.M., Darden, T.A., Kunkel, T.A. and Tomer, K.B. (2006) Mapping the Dimer Interface in the C-Terminal Domains of the Yeast MLH1-PMS1 Heterodimer. *Biochemistry*, **45**, 15458-15467.
- Dall'Antonia, F. and Schneider, T.R. (2006) SITCOM: a program for comparing sites in macromolecular substructures. *Journal of Applied Crystallography*, **39**, 618-619.
- De Rosa, M., Fasano, C., Panariello, L., Scarano, M.I., Belli, G., Iannelli, A., Ciciliano, F. and Izzo, P. (2000) Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. *Oncogene*, **19**, 1719-1723.
- de Villartay, J.P., Shimazaki, N., Charbonnier, J.B., Fischer, A., Mornon, J.P., Lieber, M.R. and Callebaut, I. (2009) A histidine in the beta-CASP domain of Artemis is critical for its full in vitro and in vivo functions. *DNA Repair (Amst)*, **8**, 202-208.
- de Vries, S.S., Baart, E.B., Dekker, M., Siezen, A., de Rooij, D.G., de Boer, P. and te Riele, H. (1999) Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev*, **13**, 523-531.
- Dessinges, M.N., Lionnet, T., Xi, X.G., Bensimon, D. and Croquette, V. (2004) Single-molecule assay reveals strand switching and enhanced processivity of UvrD. *Proc Natl Acad Sci U S A*, **101**, 6439-6444.

- Dherin, C., Gueneau, E., Francin, M., Nunez, M., Miron, S., Liberti, S.E., Rasmussen, L.J., Zinn-Justin, S., Gilquin, B., Charbonnier, J.B. and Boiteux, S. (2009) Characterization of a highly conserved binding site of Mlh1 required for exonuclease I-dependent mismatch repair. *Mol Cell Biol*, **29**, 907-918.
- Dimasi, N., Flot, D., Dupeux, F. and Marquez, J.A. (2007) Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **63**, 204-208.
- Dodson, E. (2003) Is it jolly SAD? *Acta Crystallographica Section D-Biological Crystallography*, **59**, 1958-1965.
- Downen, J.M., Putnam, C.D. and Kolodner, R.D. (2010) Functional studies and homology modeling of Msh2-Msh3 predict that mispair recognition involves DNA bending and strand separation. *Mol Cell Biol*, **30**, 3321-3328.
- Drotschmann, K., Aronshtam, A., Fritz, H.J. and Marinus, M.G. (1998) The Escherichia coli MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *Nucleic Acids Res*, **26**, 948-953.
- Duckett, D.R., Drummond, J.T., Murchie, A.I., Reardon, J.T., Sancar, A., Lilley, D.M. and Modrich, P. (1996) Human MutSalphalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. *Proc Natl Acad Sci U S A*, **93**, 6443-6447.
- Dufner, P., Marra, G., Raschle, M. and Jiricny, J. (2000) Mismatch recognition and DNA-dependent stimulation of the ATPase activity of hMutSalphalpha is abolished by a single mutation in the hMSH6 subunit. *J Biol Chem*, **275**, 36550-36555.
- Duppatla, V., Bodda, C., Urbanke, C., Friedhoff, P. and Rao, D.N. (2009) The C-terminal domain is sufficient for endonuclease activity of Neisseria gonorrhoeae MutL. *Biochem J*, **423**, 265-277.
- Durant, S.T., Morris, M.M., Illand, M., McKay, H.J., McCormick, C., Hirst, G.L., Borts, R.H. and Brown, R. (1999) Dependence on RAD52 and RAD1 for anticancer drug resistance mediated by inactivation of mismatch repair genes. *Curr Biol*, **9**, 51-54.
- Dutta, R. and Inouye, M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*, **25**, 24-28.
- Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M. and Modrich, P. (2004) A defined human system that supports bidirectional mismatch-provoked excision. *Mol Cell*, **15**, 31-41.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32**, 1792-1797.
- Ehrenstein, M.R. and Neuberger, M.S. (1999) Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *Embo J*, **18**, 3484-3490.

- Ehrenstein, M.R., Rada, C., Jones, A.M., Milstein, C. and Neuberger, M.S. (2001) Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. *Proc Natl Acad Sci U S A*, **98**, 14553-14558.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography*, **60**, 2126-2132.
- Engh, R.A. and Huber, R. (2001) *Structure quality and target parameters*. Kluwer Academic Publishers, Dordrecht.
- Engh, R.A. and Huber, R. (2006) Structure quality and target parameters. *International Tables for Crystallography*, **F**, 382-392.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U. and Sali, A. (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, **Chapter 2**, Unit 2 9.
- Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, **62**, 72-82.
- Fiorentini, P., Huang, K.N., Tishkoff, D.X., Kolodner, R.D. and Symington, L.S. (1997) Exonuclease I of *Saccharomyces cerevisiae* functions in mitotic recombination in vivo and in vitro. *Mol Cell Biol*, **17**, 2764-2773.
- Flores-Rozas, H., Clark, D. and Kolodner, R.D. (2000) Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mismatch recognition complex. *Nat Genet*, **26**, 375-378.
- Friedberg, E.C. (2003) DNA damage and repair. *Nature*, **421**, 436-440.
- Fukui, K., Nishida, M., Nakagawa, N., Masui, R. and Kuramitsu, S. (2008) Bound nucleotide controls the endonuclease activity of mismatch repair enzyme MutL. *J Biol Chem*, **283**, 12136-12145.
- Galburt, E.A. and Stoddard, B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851-13860.
- Garman, E.F. and Owen, R.L. (2006) Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography*, **62**, 32-47.
- Gellon, L., Werner, M. and Boiteux, S. (2002) Ntg2p, a *Saccharomyces cerevisiae* DNA N-glycosylase/apurinic or apyrimidinic lyase involved in base excision repair of oxidative DNA damage, interacts with the DNA mismatch repair protein Mlh1p. Identification of a Mlh1p binding motif. *J Biol Chem*, **277**, 29963-29972.
- Genschel, J., Bazemore, L.R. and Modrich, P. (2002) Human exonuclease I is required for 5' and 3' mismatch repair. *J Biol Chem*, **277**, 13302-13311.
- Genschel, J. and Modrich, P. (2003) Mechanism of 5'-directed excision in human mismatch repair. *Mol Cell*, **12**, 1077-1086.
- Georgescu, R.E., Kim, S.S., Yurieva, O., Kuriyan, J., Kong, X.P. and O'Donnell, M. (2008) Structure of a sliding clamp on DNA. *Cell*, **132**, 43-54.

- Giron-Monzon, L., Manelyte, L., Ahrends, R., Kirsch, D., Spengler, B. and Friedhoff, P. (2004) Mapping protein-protein interactions between MutL and MutH by cross-linking. *J Biol Chem*, **279**, 49338-49345.
- Glickman, B.W. and Radman, M. (1980) Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction. *Proc Natl Acad Sci U S A*, **77**, 1063-1067.
- Goldfarb, T. and Alani, E. (2005) Distinct roles for the Saccharomyces cerevisiae mismatch repair proteins in heteroduplex rejection, mismatch repair and nonhomologous tail removal. *Genetics*, **169**, 563-574.
- Green, D.W., Ingram, V.M. and Perutz, M.F. (1954) The Structure of Haemoglobin .4. Sign Determination by the Isomorphous Replacement Method. *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences*, **225**, 287-307.
- Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*, **1**, 2876-2890.
- Guanti, G., Resta, N., Simone, C., Cariola, F., Demma, I., Fiorente, P. and Gentile, M. (2000) Involvement of PTEN mutations in the genetic pathways of colorectal cancerogenesis. *Human Molecular Genetics*, **9**, 283-287.
- Guarne, A., Junop, M.S. and Yang, W. (2001) Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *Embo J*, **20**, 5521-5531.
- Guarne, A., Ramon-Maiques, S., Wolff, E.M., Ghirlando, R., Hu, X., Miller, J.H. and Yang, W. (2004) Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *Embo J*, **23**, 4134-4145.
- Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996) Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr Biol*, **6**, 1185-1187.
- Hall, M.C., Jordan, J.R. and Matson, S.W. (1998) Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD. *Embo J*, **17**, 1535-1541.
- Hall, M.C. and Matson, S.W. (1999) The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. *J Biol Chem*, **274**, 1306-1312.
- Hall, M.C., Shcherbakova, P.V., Fortune, J.M., Borchers, C.H., Dial, J.M., Tomer, K.B. and Kunkel, T.A. (2003) DNA binding by yeast Mlh1 and Pms1: implications for DNA mismatch repair. *Nucleic Acids Res*, **31**, 2025-2034.
- Hall, M.C., Shcherbakova, P.V. and Kunkel, T.A. (2002) Differential ATP binding and intrinsic ATP hydrolysis by amino-terminal domains of the yeast Mlh1 and Pms1 proteins. *J Biol Chem*, **277**, 3673-3679.
- Hall, M.C., Wang, H., Erie, D.A. and Kunkel, T.A. (2001) High affinity cooperative DNA binding by the yeast Mlh1-Pms1 heterodimer. *J Mol Biol*, **312**, 637-647.

- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*, **166**, 557-580.
- Harfe, B.D., Minesinger, B.K. and Jinks-Robertson, S. (2000) Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in the removal of frameshift intermediates in budding yeast. *Curr Biol*, **10**, 145-148.
- He, W., Zhao, Y., Zhang, C., An, L., Hu, Z., Liu, Y., Han, L., Bi, L., Xie, Z., Xue, P., Yang, F. and Hang, H. (2008) Rad9 plays an important role in DNA mismatch repair through physical interaction with MLH1. *Nucleic Acids Res*, **36**, 6406-6417.
- Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J*, **9**, 1665-1672.
- Hendriks, Y.M.C., Jagmohan-Changur, S., Van Der Klift, H.M., Morreau, H., Van Puijenbroek, M., Tops, C., Van Os, T., Wagner, A., Ausems, M.G.F.M., Gomez, E., Breuning, M.H., Brocker-Vriends, A.H.J.T., Vasen, H.F.A. and Wijnen, J.T. (2006) Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). *Gastroenterology*, **130**, 312-322.
- Hoeijmakers, J.H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature*, **411**, 366-374.
- Holdgate, G.A. (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *Biotechniques*, **31**, 164-170.
- Huang, J.Q., Kuismanen, S.A., Liu, T., Chadwick, R.B., Johnson, C.K., Stevens, M.W., Richards, S.K., Meek, J.E., Gao, X., Wright, F.A., Mecklin, J.P., Jarvinen, H.J., Gronberg, H., Bisgaard, M.L., Lindblom, A. and Peltomaki, P. (2001) MSH6 and MSH3 are rarely involved in genetic predisposition to nonpolypotic colon cancer. *Cancer Research*, **61**, 1619-1623.
- Hunter, N. and Borts, R.H. (1997) Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes Dev*, **11**, 1573-1582.
- Iino, H., Kim, K., Shimada, A., Masui, R., Kuramitsu, S. and Fukui, K. (2010) Characterization of C- and N-terminal domains of *Aquifex aeolicus* MutL endonuclease: N-terminal domain stimulates the endonuclease activity of C-terminal domain in a zinc-dependent manner. *Biosci Rep*.
- Ikenoue, T., Togo, G., Nagai, K., Ijichi, H., Kato, J., Yamaji, Y., Okamoto, M., Kato, N., Kawabe, T., Tanaka, A., Matsumura, M., Shiratori, Y. and Omata, M. (2001) Frameshift mutations at mononucleotide repeats in RAD50 recombinational DNA repair gene in colorectal cancers with microsatellite instability. *Jpn J Cancer Res*, **92**, 587-591.
- Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23-28.
- Iyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. (2006) DNA mismatch repair: functions and mechanisms. *Chem Rev*, **106**, 302-323.

- Iyer, R.R., Pohlhaus, T.J., Chen, S., Hura, G.L., Dzantiev, L., Beese, L.S. and Modrich, P. (2008) The MutS α -proliferating cell nuclear antigen interaction in human DNA mismatch repair. *J Biol Chem*, **283**, 13310-13319.
- Jacob, S. and Praz, F. (2002) DNA mismatch repair defects: role in colorectal carcinogenesis. *Biochimie*, **84**, 27-47.
- Jacobs-Palmer, E. and Hingorani, M.M. (2007) The effects of nucleotides on MutS-DNA binding kinetics clarify the role of MutS ATPase activity in mismatch repair. *J Mol Biol*, **366**, 1087-1098.
- Jager, A.C., Rasmussen, M., Bisgaard, H.C., Singh, K.K., Nielsen, F.C. and Rasmussen, L.J. (2001) HNPCC mutations in the human DNA mismatch repair gene hMLH1 influence assembly of hMutL α and hMLH1-hEXO1 complexes. *Oncogene*, **20**, 3590-3595.
- James, R.W. (1958) *The Optical Principles of the Diffraction by X-rays*. G. Bell And Sons Limited, London.
- Jang, E. and Chung, D.C. (2010) Hereditary Colon Cancer: Lynch Syndrome. *Gut and Liver*, **4**, 151-160.
- Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, **7**, 335-346.
- Jiricny, J. and Nystrom-Lahti, M. (2000) Mismatch repair defects in cancer. *Curr Opin Genet Dev*, **10**, 157-161.
- Joshi, A., Sen, S. and Rao, B.J. (2000) ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes. *Nucleic Acids Res*, **28**, 853-861.
- Junop, M.S., Obmolova, G., Rausch, K., Hsieh, P. and Yang, W. (2001) Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol Cell*, **7**, 1-12.
- Kabsch, W. (1988) Automatic Indexing of Rotation Diffraction Patterns. *J. Appl. Cryst.*, **21**, 67-71.
- Kadyrov, F.A., Dzantiev, L., Constantin, N. and Modrich, P. (2006) Endonucleolytic function of MutL α in human mismatch repair. *Cell*, **126**, 297-308.
- Kadyrov, F.A., Genschel, J., Fang, Y., Penland, E., Edelmann, W. and Modrich, P. (2009) A possible mechanism for exonuclease 1-independent eukaryotic mismatch repair. *Proc Natl Acad Sci U S A*, **106**, 8495-8500.
- Kadyrov, F.A., Holmes, S.F., Arana, M.E., Lukianova, O.A., O'Donnell, M., Kunkel, T.A. and Modrich, P. (2007) *Saccharomyces cerevisiae* MutL α is a mismatch repair endonuclease. *J Biol Chem*, **282**, 37181-37190.
- Kanao, R., Hanaoka, F. and Masutani, C. (2009) A novel interaction between human DNA polymerase η and MutL α . *Biochem Biophys Res Commun*, **389**, 40-45.

- Kantardjieff, K.A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, **12**, 1865-1871.
- Karle, J. and Hauptman, H. (1956) A Theory of Phase Determination for the 4 Types of Non-Centrosymmetric Space Groups 1P222, 2P22, 3P12, 3P22. *Acta Crystallographica*, **9**, 635-651.
- Kelly, S.M., Jess, T.J. and Price, N.C. (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta*, **1751**, 119-139.
- Kim, T.G., Cha, H.J., Lee, H.J., Heo, S.D., Choi, K.Y., Ku, J.K. and Ban, C. (2009) Structural insights of the nucleotide-dependent conformational changes of *Thermotoga maritima* MutL using small-angle X-ray scattering analysis. *J Biochem*, **145**, 199-206.
- Kleczkowska, H.E., Marra, G., Lettieri, T. and Jiricny, J. (2001) hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev*, **15**, 724-736.
- Kneitz, B., Cohen, P.E., Avdievich, E., Zhu, L., Kane, M.F., Hou, H., Jr., Kolodner, R.D., Kucherlapati, R., Pollard, J.W. and Edelmann, W. (2000) MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes Dev*, **14**, 1085-1097.
- Kolas, N.K., Svetlanov, A., Lenzi, M.L., Macaluso, F.P., Lipkin, S.M., Liskay, R.M., Greally, J., Edelmann, W. and Cohen, P.E. (2005) Localization of MMR proteins on meiotic chromosomes in mice indicates distinct functions during prophase I. *J Cell Biol*, **171**, 447-458.
- Kondo, E., Horii, A. and Fukushige, S. (2001) The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res*, **29**, 1695-1702.
- Korhonen, M.K., Vuorenmaa, E. and Nystrom, M. (2008) The first functional study of MLH3 mutations found in cancer patients. *Genes Chromosomes Cancer*, **47**, 803-809.
- Kosinski, J., Hinrichsen, I., Bujnicki, J.M., Friedhoff, P. and Plotz, G. (2010) Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Hum Mutat*, **31**, 975-982.
- Kosinski, J., Plotz, G., Guarne, A., Bujnicki, J.M. and Friedhoff, P. (2008) The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. *J Mol Biol*, **382**, 610-627.
- Kosinski, J., Steindorf, I., Bujnicki, J.M., Giron-Monzon, L. and Friedhoff, P. (2005) Analysis of the Quaternary Structure of the MutL C-terminal Domain. *Journal of Molecular Biology*, **351**, 895-909.
- Kramer, B., Kramer, W. and Fritz, H.J. (1984) Different Base Base Mismatches Are Corrected with Different Efficiencies by the Methyl-Directed DNA Mismatch-Repair System of *Escherichia-Coli*. *Cell*, **38**, 879-887.
- Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233-1243.

- Kuismanen, S.A., Holmberg, M.T., Salovaara, R., de la Chapelle, A. and Peltomäki, P. (2000) Genetic and epigenetic modification of MLH1 accounts for a major share of microsatellite-unstable colorectal cancers. *American Journal of Pathology*, **156**, 1773-1779.
- Kunkel, T.A. (2004) DNA replication fidelity. *J Biol Chem*, **279**, 16895-16898.
- Kunkel, T.A. and Erie, D.A. (2005) DNA mismatch repair. *Annu Rev Biochem*, **74**, 681-710.
- Kunz, C., Saito, Y. and Schar, P. (2009) DNA Repair in mammalian cells: Mismatched repair: variations on a theme. *Cell Mol Life Sci*, **66**, 1021-1038.
- Laguri, C., Duband-Goulet, I., Friedrich, N., Axt, M., Belin, P., Callebaut, I., Gilquin, B., Zinn-Justin, S. and Couprie, J. (2008) Human mismatch repair protein MSH6 contains a PWWP domain that targets double stranded DNA. *Biochemistry*, **47**, 6199-6207.
- Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H., de Wind, N. and Sixma, T.K. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature*, **407**, 711-717.
- Langland, G., Kordich, J., Creaney, J., Goss, K.H., Lillard-Wetherell, K., Bebenek, K., Kunkel, T.A. and Groden, J. (2001) The Bloom's syndrome protein (BLM) interacts with MLH1 but is not required for DNA mismatch repair. *J Biol Chem*, **276**, 30031-30035.
- Larrea, A.A., Lujan, S.A. and Kunkel, T.A. (2010) SnapShot: DNA mismatch repair. *Cell*, **141**, 730 e731.
- Lee, J.Y., Chang, J., Joseph, N., Ghirlando, R., Rao, D.N. and Yang, W. (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. *Mol Cell*, **20**, 155-166.
- Lee, S.D. and Alani, E. (2006) Analysis of interactions between mismatch repair initiation factors and the replication processivity factor PCNA. *J Mol Biol*, **355**, 175-184.
- Li, F., Liu, Q., Chen, Y.Y., Yu, Z.N., Zhang, Z.P., Zhou, Y.F., Deng, J.Y., Bi, L.J. and Zhang, X.E. (2008) Escherichia coli mismatch repair protein MutL interacts with the clamp loader subunits of DNA polymerase III. *Mutat Res*, **637**, 101-110.
- Li, G.M. (1999) The role of mismatch repair in DNA damage-induced apoptosis. *Oncol Res*, **11**, 393-400.
- Li, G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res*, **18**, 85-98.
- Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D. and Scharff, M.D. (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev*, **18**, 1-11.
- Liberti, S.E., Andersen, S.D., Wang, J., May, A., Miron, S., Perderiset, M., Keijzers, G., Nielsen, F.C., Charbonnier, J.B., Bohr, V.A. and Rasmussen, L.J. (2010) Bi-directional routing of DNA mismatch repair protein human exonuclease 1 to replication foci and DNA double strand breaks. *DNA Repair (Amst)*, **10**, 73-86.

- Lindblom, A., Tannergard, P., Werelius, B. and Nordenskjold, M. (1993) Genetic-Mapping of a 2Nd Locus Predisposing to Hereditary Nonpolyposis Colon-Cancer. *Nature Genetics*, **5**, 279-282.
- Lipkin, S.M., Moens, P.B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J.W., Green, E.D., Schwartzberg, P., Collins, F.S. and Cohen, P.E. (2002) Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat Genet*, **31**, 385-390.
- Liu, T., Yan, H., Kuismanen, S., Percesepe, A., Bisgaard, M.L., Pedroni, M., Benatti, P., Kinzler, K.W., Vogelstein, B., de Leon, M.P., Peltomaki, P. and Lindblom, A. (2001) The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. *Cancer Research*, **61**, 7798-7802.
- Lo Conte, L., Chothia, C. and Janin, J. (1999) The atomic structure of protein-protein recognition sites. *J Mol Biol*, **285**, 2177-2198.
- Loeb, L.A. (1994) Microsatellite Instability - Marker of a Mutator Phenotype in Cancer. *Cancer Research*, **54**, 5059-5063.
- Loukola, A., Vilkki, S., Singh, J., Launonen, V. and Aaltonen, L.A. (2000) Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer. *American Journal of Pathology*, **157**, 347-352.
- Lu, A.L., Clark, S. and Modrich, P. (1983) Methyl-directed repair of DNA base-pair mismatches in vitro. *Proc Natl Acad Sci U S A*, **80**, 4639-4643.
- Lynch, H.T. and de la Chapelle, A. (2003) Genomic medicine - Hereditary colorectal cancer. *New England Journal of Medicine*, **348**, 919-932.
- Lynch, H.T., Shaw, M.W., Magnuson, C.W., Larsen, A.L. and Krush, A.J. (1966) Hereditary Factors in Cancer - Study of 2 Large Midwestern Kindreds. *Archives of Internal Medicine*, **117**, 206-&.
- Malivert, L., Ropars, V., Nunez, M., Drevet, P., Miron, S., Faure, G., Guerois, R., Mornon, J.P., Revy, P., Charbonnier, J.B., Callebaut, I. and de Villartay, J.P. (2010) Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem*, **285**, 26475-26483.
- Malkhosyan, S., Rampino, N., Yamamoto, H. and Perucho, M. (1996) Frameshift mutator mutations. *Nature*, **382**, 499-500.
- Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L.Z., Lutterbaugh, J., Fan, R.S., Zborowska, E., Kinzler, K.W., Vogelstein, B., Brattain, M. and Willson, J.K.V. (1995) Inactivation of the Type-II Tgf-Beta Receptor in Colon-Cancer Cells with Microsatellite Instability. *Science*, **268**, 1336-1338.
- Marsischky, G.T., Filosi, N., Kane, M.F. and Kolodner, R. (1996) Redundancy of *Saccharomyces cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev*, **10**, 407-420.
- Marsischky, G.T. and Kolodner, R.D. (1999) Biochemical characterization of the interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 complex and mispaired bases in DNA. *J Biol Chem*, **274**, 26668-26682.

- Martik, D., Baitinger, C. and Modrich, P. (2004) Differential specificities and simultaneous occupancy of human MutS α nucleotide binding sites. *J Biol Chem*, **279**, 28402-28410.
- Martomo, S.A., Yang, W.W. and Gearhart, P.J. (2004) A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *J Exp Med*, **200**, 61-68.
- Matson, S.W. and Robertson, A.B. (2006) The UvrD helicase and its modulation by the mismatch repair protein MutL. *Nucleic Acids Res*, **34**, 4089-4097.
- Mauris, J. and Evans, T.C. (2009) Adenosine triphosphate stimulates Aquifex aeolicus MutL endonuclease activity. *PLoS One*, **4**, e7175.
- Mauris, J. and Evans, T.C., Jr. (2010) A human PMS2 homologue from Aquifex aeolicus stimulates an ATP-dependent DNA helicase. *J Biol Chem*, **285**, 11087-11092.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
- McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404-405.
- McNally, R., Bowman, G.D., Goedken, E.R., O'Donnell, M. and Kuriyan, J. (2010) Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC Struct Biol*, **10**, 3.
- Mechanic, L.E., Frankel, B.A. and Matson, S.W. (2000) Escherichia coli MutL loads DNA helicase II onto DNA. *J Biol Chem*, **275**, 38337-38346.
- Mendillo, M.L., Putnam, C.D., Mo, A.O., Jamison, J.W., Li, S., Woods, V.L., Jr. and Kolodner, R.D. (2010) Probing DNA- and ATP-mediated conformational changes in the MutS family of mispair recognition proteins using deuterium exchange mass spectrometry. *J Biol Chem*, **285**, 13170-13182.
- Miyaki, M., Konishi, M., Tanaka, K., KikuchiYanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M. and Mori, T. (1997) Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature Genetics*, **17**, 271-272.
- Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem*, **65**, 101-133.
- Mukherjee, S., Law, S.M. and Feig, M. (2009) Deciphering the mismatch recognition cycle in MutS and MSH2-MSH6 using normal-mode analysis. *Biophys J*, **96**, 1707-1720.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography*, **53**, 240-255.
- Namadurai, S., Jain, D., Kulkarni, D.S., Tabib, C.R., Friedhoff, P., Rao, D.N. and Nair, D.T. (2010) The C-terminal domain of the MutL homolog from Neisseria gonorrhoeae forms an inverted homodimer. *PLoS One*, **5**, e13726.

- Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H., Perrakis, A. and Sixma, T.K. (2003) Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res*, **31**, 4814-4821.
- Nishant, K.T., Plys, A.J. and Alani, E. (2008) A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in Saccharomyces cerevisiae. *Genetics*, **179**, 747-755.
- Obmolova, G., Ban, C., Hsieh, P. and Yang, W. (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature*, **407**, 703-710.
- Ou, J., Rasmussen, M., Westers, H., Andersen, S.D., Jager, P.O., Kooi, K.A., Niessen, R.C., Eggen, B.J., Nielsen, F.C., Kleibeuker, J.H., Sijmons, R.H., Rasmussen, L.J. and Hofstra, R.M. (2009) Biochemical characterization of MLH3 missense mutations does not reveal an apparent role of MLH3 in Lynch syndrome. *Genes Chromosomes Cancer*, **48**, 340-350.
- Painter, J. and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D-Biological Crystallography*, **62**, 439-450.
- Pang, Q., Prolla, T.A. and Liskay, R.M. (1997) Functional domains of the Saccharomyces cerevisiae Mlh1p and Pms1p DNA mismatch repair proteins and their relevance to human hereditary nonpolyposis colorectal cancer-associated mutations. *Mol Cell Biol*, **17**, 4465-4473.
- Park, J., Jeon, Y., In, D., Fishel, R., Ban, C. and Lee, J.B. (2010) Single-Molecule Analysis Reveals the Kinetics and Physiological Relevance of MutL-ssDNA Binding. *PLoS One*, **5**, e15496.
- Parker, B.O. and Marinus, M.G. (1992) Repair of DNA Heteroduplexes Containing Small Heterologous Sequences in Escherichia-Coli. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 1730-1734.
- Pedrazzi, G., Perrera, C., Blaser, H., Kuster, P., Marra, G., Davies, S.L., Ryu, G.H., Freire, R., Hickson, I.D., Jiricny, J. and Stagljar, I. (2001) Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1. *Nucleic Acids Res*, **29**, 4378-4386.
- Peltomaki, P. (2001) Deficient DNA mismatch repair: a common etiologic factor for colon cancer. *Hum Mol Genet*, **10**, 735-740.
- Peltomaki, P., Aaltonen, L.A., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Green, J.S., Jass, J.R., Weber, J.L., Leach, F.S., Petersen, G.M., Hamilton, S.R., Delachapelle, A. and Vogelstein, B. (1993) Genetic-Mapping of a Locus Predisposing to Human Colorectal-Cancer. *Science*, **260**, 810-812.
- Peltomaki, P. and Vasen, H. (2004) Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database. *Dis Markers*, **20**, 269-276.
- Peng, M., Litman, R., Xie, J., Sharma, S., Brosh, R.M., Jr. and Cantor, S.B. (2007) The FANCI/MutLalpha interaction is required for correction of the cross-link response in FA-J cells. *Embo J*, **26**, 3238-3249.

- Phung, Q.H., Winter, D.B., Alrefai, R. and Gearhart, P.J. (1999) Hypermutation in Ig V genes from mice deficient in the MLH1 mismatch repair protein. *J Immunol*, **162**, 3121-3124.
- Pierce, A.J., Stark, J.M., Araujo, F.D., Moynahan, M.E., Berwick, M. and Jasin, M. (2001) Double-strand breaks and tumorigenesis. *Trends Cell Biol*, **11**, S52-59.
- Pillon, M.C., Lorenowicz, J.J., Uckelmann, M., Klocko, A.D., Mitchell, R.R., Chung, Y.S., Modrich, P., Walker, G.C., Simmons, L.A., Friedhoff, P. and Guarne, A. (2010a) Structure of the endonuclease domain of MutL: unlicensed to cut. *Mol Cell*, **39**, 145-151.
- Pillon, M.C., Miller, J.H. and Guarne, A. (2010b) The endonuclease domain of MutL interacts with the beta sliding clamp. *DNA Repair (Amst)*.
- Plotz, G., ... and Zeuzem, S. (2003) N-terminus of hMLH1 confers interaction of hMutLa and hMutb with hMutSa. *Nuc. acid Res*, **31**, 3217-3226.
- Plotz, G., Welsch, C., Giron-Monzon, L., Friedhoff, P., Albrecht, M., Piiper, A., Biondi, R.M., Lengauer, T., Zeuzem, S. and Raedle, J. (2006) Mutations in the MutSalph interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Res*, **34**, 6574-6586.
- Pluciennik, A., Burdett, V., Lukianova, O., O'Donnell, M. and Modrich, P. (2009) Involvement of the beta clamp in methyl-directed mismatch repair in vitro. *J Biol Chem*, **284**, 32782-32791.
- Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A. and Modrich, P. (2010) PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc Natl Acad Sci U S A*, **107**, 16066-16071.
- Pluciennik, A. and Modrich, P. (2007) Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc Natl Acad Sci U S A*, **104**, 12709-12713.
- Podust, L.M., Podust, V.N., Sogo, J.M. and Hubscher, U. (1995) Mammalian DNA polymerase auxiliary proteins: analysis of replication factor C-catalyzed proliferating cell nuclear antigen loading onto circular double-stranded DNA. *Mol Cell Biol*, **15**, 3072-3081.
- Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P. and Craescu, C.T. (2003) Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin. *J Biol Chem*, **278**, 40252-40261.
- Potocnik, U., Glavac, D., Golouh, R. and Ravnik-Glavac, M. (2001) Causes of microsatellite instability in colorectal tumors: implications for hereditary non-polyposis colorectal cancer screening. *Cancer Genetics and Cytogenetics*, **126**, 85-96.
- Poulogiannis, G., Frayling, I.M. and Arends, M.J. (2010) DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology*, **56**, 167-179.
- Prolla, T.A., Baker, S.M., Harris, A.C., Tsao, J.L., Yao, X., Bronner, C.E., Zheng, B., Gordon, M., Reneker, J., Arnheim, N., Shibata, D., Bradley, A. and Liskay, R.M. (1998) Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. *Nat Genet*, **18**, 276-279.

- Qiu, J., Qian, Y., Chen, V., Guan, M.X. and Shen, B. (1999) Human exonuclease 1 functionally complements its yeast homologues in DNA recombination, RNA primer removal, and mutation avoidance. *J Biol Chem*, **274**, 17893-17900.
- Rada, C., Ehrenstein, M.R., Neuberger, M.S. and Milstein, C. (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*, **9**, 135-141.
- Rajewsky, K., Forster, I. and Cumano, A. (1987) Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science*, **238**, 1088-1094.
- Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C. and Perucho, M. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*, **275**, 967-969.
- Raschle, M., Dufner, P., Marra, G. and Jiricny, J. (2002) Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J Biol Chem*, **277**, 21810-21820.
- Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P. and Jiricny, J. (1999) Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J Biol Chem*, **274**, 32368-32375.
- Rasmussen, L.J., Rasmussen, M., Lee, B., Rasmussen, A.K., Wilson, D.M., 3rd, Nielsen, F.C. and Bisgaard, H.C. (2000) Identification of factors interacting with hMSH2 in the fetal liver utilizing the yeast two-hybrid system. In vivo interaction through the C-terminal domains of hEXO1 and hMSH2 and comparative expression analysis. *Mutat Res*, **460**, 41-52.
- Robertson, A., Pattishall, S.R. and Matson, S.W. (2006) The DNA binding activity of MutL is required for methyl-directed mismatch repair in Escherichia coli. *J Biol Chem*, **281**, 8399-8408.
- Rossmann, M.G. (1990) The molecular replacement method. *Acta Crystallogr A*, **46** (Pt 2), 73-82.
- Rould, M.A. (1997) Screening for heavy-atom derivatives and obtaining accurate isomorphous differences. *Macromolecular Crystallography, Pt A*, **276**, 461-472.
- Sacho, E.J., Kadyrov, F.A., Modrich, P., Kunkel, T.A. and Erie, D.A. (2008) Direct visualization of asymmetric adenine-nucleotide-induced conformational changes in MutL alpha. *Mol Cell*, **29**, 112-121.
- Santucci-Darmanin, S., Neyton, S., Lespinasse, F., Saunieres, A., Gaudray, P. and Paquis-Flucklinger, V. (2002) The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Hum Mol Genet*, **11**, 1697-1706.
- Santucci-Darmanin, S. and Paquis-Flucklinger, V. (2003) [Homologs of MutS and MutL during mammalian meiosis]. *Med Sci (Paris)*, **19**, 85-91.
- Schanz, S., Castor, D., Fischer, F. and Jiricny, J. (2009) Interference of mismatch and base excision repair during the processing of adjacent U/G mispairs may play a key role in somatic hypermutation. *Proc Natl Acad Sci U S A*, **106**, 5593-5598.

- Schmutte, C., Sadoff, M.M., Shim, K.S., Acharya, S. and Fishel, R. (2001) The interaction of DNA mismatch repair proteins with human exonuclease I. *J Biol Chem*, **276**, 33011-33018.
- Schofield, M.J., Nayak, S., Scott, T.H., Du, C. and Hsieh, P. (2001) Interaction of Escherichia coli MutS and MutL at a DNA mismatch. *J Biol Chem*, **276**, 28291-28299.
- Schrader, C.E., Edelmann, W., Kucherlapati, R. and Stavnezer, J. (1999) Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *J Exp Med*, **190**, 323-330.
- Schrader, C.E., Guikema, J.E., Linehan, E.K., Selsing, E. and Stavnezer, J. (2007) Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *J Immunol*, **179**, 6064-6071.
- Schwartz, S., Yamamoto, H., Navarro, M., Maestro, M., Reventos, J. and Perucho, M. (1999) Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype. *Cancer Research*, **59**, 2995-3002.
- Selmane, T., Schofield, M.J., Nayak, S., Du, C. and Hsieh, P. (2003) Formation of a DNA mismatch repair complex mediated by ATP. *J Mol Biol*, **334**, 949-965.
- Shcherbakova, P.V., Hall, M.C., Lewis, M.S., Bennett, S.E., Martin, K.J., Bushel, P.R., Afshari, C.A. and Kunkel, T.A. (2001) Inactivation of DNA mismatch repair by increased expression of yeast MLH1. *Mol Cell Biol*, **21**, 940-951.
- Shell, S.S., Putnam, C.D. and Kolodner, R.D. (2007) Chimeric *Saccharomyces cerevisiae* Msh6 protein with an Msh3 mispair-binding domain combines properties of both proteins. *Proc Natl Acad Sci U S A*, **104**, 10956-10961.
- Shimada, A., Masui, R., Nakagawa, N., Takahata, Y., Kim, K., Kuramitsu, S. and Fukui, K. (2010) A novel single-stranded DNA-specific 3'-5' exonuclease, *Thermus thermophilus* exonuclease I, is involved in several DNA repair pathways. *Nucleic Acids Res*, **38**, 5692-5705.
- Simmons, L.A., Davies, B.W., Grossman, A.D. and Walker, G.C. (2008) Beta clamp directs localization of mismatch repair in *Bacillus subtilis*. *Mol Cell*, **29**, 291-301.
- Smogorzewska, A., Desetty, R., Saito, T.T., Schlabach, M., Lach, F.P., Sowa, M.E., Clark, A.B., Kunkel, T.A., Harper, J.W., Colaiacovo, M.P. and Elledge, S.J. (2010) A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell*, **39**, 36-47.
- Snowden, T., Acharya, S., Butz, C., Berardini, M. and Fishel, R. (2004) hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Mol Cell*, **15**, 437-451.
- Stites, W.E. (1997) Protein-protein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem Rev*, **97**, 1233-1250.
- Suhre, K. and Sanejouand, Y.H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*, **32**, W610-614.

- Swann, P.F., Waters, T.R., Moulton, D.C., Xu, Y.Z., Zheng, Q., Edwards, M. and Mace, R. (1996) Role of postreplicative DNA mismatch repair in the cytotoxic action of thioguanine. *Science*, **273**, 1109-1111.
- Szankasi, P. and Smith, G.R. (1992) A DNA exonuclease induced during meiosis of *Schizosaccharomyces pombe*. *J Biol Chem*, **267**, 3014-3023.
- Szankasi, P. and Smith, G.R. (1995) A role for exonuclease I from *S. pombe* in mutation avoidance and mismatch correction. *Science*, **267**, 1166-1169.
- Taylor, G. (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr*, **59**, 1881-1890.
- Thakur, A.S., Robin, G., Guncar, G., Saunders, N.F., Newman, J., Martin, J.L. and Kobe, B. (2007) Improved success of sparse matrix protein crystallization screening with heterogeneous nucleating agents. *PLoS ONE*, **2**, e1091.
- Thomas, D.C., Roberts, J.D. and Kunkel, T.A. (1991) Heteroduplex repair in extracts of human HeLa cells. *J Biol Chem*, **266**, 3744-3751.
- Tishkoff, D.X., Boerger, A.L., Bertrand, P., Filosi, N., Gaida, G.M., Kane, M.F. and Kolodner, R.D. (1997) Identification and characterization of *Saccharomyces cerevisiae* EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc Natl Acad Sci U S A*, **94**, 7487-7492.
- Tomer, G., Buermeier, A.B., Nguyen, M.M. and Liskay, R.M. (2002) Contribution of human mlh1 and pms2 ATPase activities to DNA mismatch repair. *J Biol Chem*, **277**, 21801-21809.
- Tran, P.T. and Liskay, R.M. (2000) Functional studies on the candidate ATPase domains of *Saccharomyces cerevisiae* MutLalpha. *Mol Cell Biol*, **20**, 6390-6398.
- Tran, P.T., Simon, J.A. and Liskay, R.M. (2001) Interactions of Exo1p with components of MutLalpha in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **98**, 9760-9765.
- Umar, A., Buermeier, A.B., Simon, J.A., Thomas, D.C., Clark, A.B., Liskay, R.M. and Kunkel, T.A. (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell*, **87**, 65-73.
- Vagin, A. and Teplyakov, A. (2009) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr*, **66**, 22-25.
- Wagner, R., Jr. and Meselson, M. (1976) Repair tracts in mismatched DNA heteroduplexes. *Proc Natl Acad Sci U S A*, **73**, 4135-4139.
- Wang, H., Yang, Y., Schofield, M.J., Du, C.W., Fridman, Y., Lee, S.D., Larson, E.D., Drummond, J.T., Alani, E., Hsieh, P. and Erie, D.A. (2003) DNA bending and unbending by MutS govern mismatch recognition and specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 14822-14827.
- Wang, T.F., Kleckner, N. and Hunter, N. (1999) Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proc Natl Acad Sci U S A*, **96**, 13914-13919.

- Wang, Y. and Qin, J. (2003) MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proc Natl Acad Sci U S A*, **100**, 15387-15392.
- Warren, J.J., Pohlhaus, T.J., Changela, A., Iyer, R.R., Modrich, P.L. and Beese, L.S. (2007) Structure of the human MutS α DNA lesion recognition complex. *Mol Cell*, **26**, 579-592.
- Warthin, A.S. (1913) Heredity with Reference to Carcinoma. *Arch. Intern. Med.*, **12**, 546-555.
- Wei, K., Clark, A.B., Wong, E., Kane, M.F., Mazur, D.J., Parris, T., Kolas, N.K., Russell, R., Hou, H., Jr., Kneitz, B., Yang, G., Kunkel, T.A., Kolodner, R.D., Cohen, P.E. and Edelman, W. (2003) Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes Dev*, **17**, 603-614.
- Wei, K., Kucherlapati, R. and Edelman, W. (2002) Mouse models for human DNA mismatch-repair gene defects. *Trends in Molecular Medicine*, **8**, 346-353.
- Welsh, K.M., Lu, A.L., Clark, S. and Modrich, P. (1987) Isolation and characterization of the *Escherichia coli* mutH gene product. *J Biol Chem*, **262**, 15624-15629.
- Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R. and Gearhart, P.J. (2005) MSH2-MSH6 stimulates DNA polymerase ϵ , suggesting a role for A:T mutations in antibody genes. *J Exp Med*, **201**, 637-645.
- Winter, D.B., Phung, Q.H., Umar, A., Baker, S.M., Tarone, R.E., Tanaka, K., Liskay, R.M., Kunkel, T.A., Bohr, V.A. and Gearhart, P.J. (1998) Altered spectra of hypermutation in antibodies from mice deficient for the DNA mismatch repair protein PMS2. *Proc Natl Acad Sci U S A*, **95**, 6953-6958.
- Woldbye, F. and Bagger, S. (1963) Optical Rotatory Dispersion .2. Effect of Circular Dichroism Upon Measurements of Optical Rotatory Dispersion by Means of a Spectrophotometer with a Polarimetric Attachment - a New Principle for Measuring Circular Dichroism. *Acta Chemica Scandinavica*, **17**, 817-&.
- Woods, M.O., Williams, P., Careen, A., Edwards, L., Bartlett, S., McLaughlin, J.R. and Younghusband, H.B. (2007) A new variant database for mismatch repair genes associated with Lynch syndrome. *Human Mutation*, **28**, 669-673.
- Wu, X., Tsai, C.Y., Patam, M.B., Zan, H., Chen, J.P., Lipkin, S.M. and Casali, P. (2006) A role for the MutL mismatch repair Mlh3 protein in immunoglobulin class switch DNA recombination and somatic hypermutation. *J Immunol*, **176**, 5426-5437.
- Xie, J., Litman, R., Wang, S., Peng, M., Guillemette, S., Rooney, T. and Cantor, S.B. (2010) Targeting the FANCDJ-BRCA1 interaction promotes a switch from recombination to poleta-dependent bypass. *Oncogene*, **29**, 2499-2508.
- Yamaguchi, M., Dao, V. and Modrich, P. (1998) MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *J Biol Chem*, **273**, 9197-9201.
- Yan, T., Berry, S.E., Desai, A.B. and Kinsella, T.J. (2003) DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells. *Clin Cancer Res*, **9**, 2327-2334.

- Yang, W. (2006) Human MutLalpha: The jack of all trades in MMR is also an endonuclease. *DNA Repair (Amst)*.
- Yao, X., Buermeier, A.B., Narayanan, L., Tran, D., Baker, S.M., Prolla, T.A., Glazer, P.M., Liskay, R.M. and Arnheim, N. (1999) Different mutator phenotypes in Mlh1- versus Pms2-deficient mice. *Proc Natl Acad Sci U S A*, **96**, 6850-6855.
- York, S.J. and Modrich, P. (2006) Mismatch repair-dependent iterative excision at irreparable O6-methylguanine lesions in human nuclear extracts. *J Biol Chem*, **281**, 22674-22683.
- Yoshioka, K., Yoshioka, Y. and Hsieh, P. (2006) ATR kinase activation mediated by MutSalpha and MutLalpha in response to cytotoxic O6-methylguanine adducts. *Mol Cell*, **22**, 501-510.
- Zakharyevich, K., Ma, Y., Tang, S., Hwang, P.Y., Boiteux, S. and Hunter, N. (2010) Temporally and biochemically distinct activities of Exo1 during meiosis: double-strand break resection and resolution of double holliday junctions. *Mol Cell*, **40**, 1001-1015.
- Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.
- Zhang, Y., Yuan, F., Presnell, S.R., Tian, K., Gao, Y., Tomkinson, A.E., Gu, L. and Li, G.M. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell*, **122**, 693-705.

ANNEXE

ANNEXE 1: PREPARATION DE BACTERIES CHIMIOCOMPETENTES (INOUE ET AL., 1990)

A- LA CULTURE BACTERIENNE

- 1- A partir de la souche congelée à -80°C , réaliser un étalement sur un milieu SOB agar. Incuber à 37°C pendant 16 à 24h.
- 2- Le lendemain soir, ensemercer 100 ml de SOB liquide préchauffé à 25°C avec quelques colonies (4-5) piquées sur la boîte et incuber à 25°C jusqu'au lendemain matin.
- 3- Le matin, mesurer l'absorbance à 600 nm qui ne doit pas excéder 0.5. Les étapes 2-3 sont délicates car un ensemencement trop fort la veille au soir conduit à une absorbance trop forte le matin. De plus la vitesse de croissance varie d'une souche à l'autre. Elle peut nécessiter une mise au point. On peut éventuellement ensemercer de manières différentes (1 clones, 2 clones, 3 clones, ...) plusieurs milieux SOB et récupérer le matin celui qui a une absorbance inférieure à 0.5.
- 4- Poursuivre éventuellement la culture jusqu'à $A_{600} = 0.5$.
- 5- Mettre la culture dans des Falcons de 50 ml pré-refroidis dans la glace et laisser dans la glace 10minutes

B- LE TRAITEMENT CHIMIQUE

A PARTIR DE MAINTENANT, TOUJOURS TRAVAILLER DANS LA GLACE

La centrifugeuse doit être refroidie à l'avance et le tampon doit être froid, refroidir cônes, pipettes à -20°C et travailler en chambre froide. Ne pas tenir les tubes dans la main par le fond (réchauffement des cellules)

- 1- Centrifuger la culture 20 min à 2500g. Eliminer le surnageant par retournement et recentrifuger les tubes quelques secondes. Eliminer alors les traces de surnageant avec une pipette mécanique de type P200.
Remarque : des bactéries du culot peuvent se resuspendre légèrement : les éliminer avec le surnageant.
- 2- Re-suspendre le culot **délicatement** par aspiration refoulement avec 33 ml **en tout** de tampon TB froid **pour les 2 falcons (soit 16,5ml par falcon)**. Il ne doit pas rester d'agrégats de cellules en suspension. Incuber 10 min. dans la glace. (note: ne pas mettre de Javel trop concentrée avec le TB, réaction violente)
- 3- Centrifuger la culture 10 min à 2500g. Eliminer le surnageant par retournement et recentrifuger les tubes quelques secondes. Eliminer alors les traces de surnageant avec une pipette mécanique de type P200.

Remarque : des bactéries du culot peuvent se re-suspendre légèrement : les éliminer avec le surnageant.

4- Resuspendre le culot **délicatement** par aspiration refoulement avec 8 ml **en tout** de tampon TB froid **pour les 2 falcons (soit 4ml par falcon)**. Il ne doit pas rester d'agrégats de cellules en suspension.

5- Ajouter 600 µl de DMSO (7% final) goutte-à-goutte en remuant régulièrement. Le DMSO cristallisant au froid, préparer à l'avance à température ambiante dans un tube eppendorf 600 µl de DMSO et venir en chambre froide pour l'ajouter au dernier moment. Laisser 10 minutes dans la glace.

6- Aliquoter en fraction de 100 µl dans des tubes stériles pré-refroidis dans un bac à glace (tubes préparés à l'avance). Immédiatement congeler dans l'azote liquide. Ces tubes sont ensuite rangés dans une boîte de congélation en carton comportant une étiquette (nom souche, initiales personne, date) refroidie à -80°C . Cette boîte reste stockée à -80°C .

7- Mesurer la compétence des cellules. Voir ci-dessous. Reporter la compétence sur le classeur commun des souches et/ou sur la boîte.

C- TRANSFORMATION- MESURE DE COMPETENCE (compétence attendue 10^8)

1- Décongeler un ou plusieurs aliquots de bactéries chimiocompétentes dans la glace. Compter 50 µl par transformation. Laisser décongeler (environ 15 minutes). Mettre 1 µl de plasmide préparé par Midiprep ou Mini-Prep (env 100 ng) dans un falcon 14 ml se trouvant déjà dans la glace. Ajouter 50 µl de cellules compétentes décongelées et laisser 30 mn dans la glace. Il est conseillé d'utiliser des cônes refroidis 15 mn à -20°C pour pipeter les cellules compétentes. Pour un calcul de compétence, utiliser des quantités plus faibles de pUC18 (0,1 pg, 1 pg, 10 pg et 100 pg)

2- Réaliser un choc thermique en plaçant les échantillons 30 s à 42°C , puis remettre les tubes dans la glace 2 mn.

3- Ajouter 250 µl de milieu SOC préchauffé à 42°C dans le tube falcon 14 ml contenant les cellules compétentes transformées .

4- Incuber 45 mn à 1 h max à 37°C , 250 rpm.

5- A partir de là, on peut :

soit étaler 50 µl et 250 µl des bactéries transformées sur une boîte LB agar + antibiotiques appropriés

soit ensemer une préculture plus grande dans 3 ml à 50 ml de LB + antibio appropriés, préchauffé à 37°C .

6- Incuber une nuit à 37°C .

Remarques

1) Calcul de la compétence des cellules : diluer en cascade le plasmide de référence (ex pUC18) jusqu'à des concentrations de 100pg.µl⁻¹ , 10pg.µl⁻¹.et 1pg.µl⁻¹ . Pour diluer le plasmide, effectuer des petites dilutions successives dans 20ul d'eau ppi.

Calcul Compétence = nbr unités formant colonies (CFU)/ µg DNA introduit

Ex : 100 colonies avec 1pg de DNA

$$\text{Compétence} = 100 / 1 \cdot 10^{-6} = 10^8$$

2) Valeur de compétence obtenue. Ce protocole a donné des compétences récemment de 4 10⁸ avec des cellules XL1Blue (Pascal Belin). On doit obtenir le même type de valeur avec des DH5 . En cas de valeur inférieure à 10⁸, il y a probablement un problème à une étape.

Avec les cellules pour l'expression BL21, BL21 star, etc, on doit obtenir des valeurs plus faibles autour de 10⁶.

D- PREPARATION DES DIFFERENTES SOLUTIONS

Tous les produits utilisés doivent être le plus purs possible et récents!! Les solutions sont toutes faites avec de l'eau milliQ **même les milieux de culture!!**

Milieu SOB : 2% bacto tryptone (Difco), 0,5% yeast extract (Difco), 10mM NaCl (Sigma), 2,5mM KCl (Sigma)

Autoclaver 120°C 20'

ajouter 10mM MgCl₂ (Sigma), 10mM MgSO₄ (Sigma) (solutions 1M filtrées 0,22µm ajoutées à 10mL/L)

Milieu SOC : SOB + 20mM glucose (Sigma)

Tampon TB : Pour 500mL : PIPES 1,51 g (Fluka; 10 mM final) ou HEPES 1,19g (Sigma; 10mM final) ; 1,10g CaCl₂ (Sigma; 15mM final) ; 9,32g KCl (Sigma; 250mM final)

Ajuster à pH 6,7 avec KOH

5,44g MnCl₂ (Sigma; 55mM final)

filtrer 0,22µm et stocker à 4°C

ANNEXE 2: COMPOSITION DU MILIEU DE CULTURE POUR LA PRODUCTION DE PROTEINE MARQUEE AU SELENIUM

Protocole fourni par K. Blondeau (IBBMC)

Sels

Na ₂ HPO ₄	8g/l _{culture}
KH ₂ PO ₄	2g/l _{culture}
NaCl	0.5g/l _{culture}
MgSO ₄	0.25g/l _{culture}
CaCl ₂	0.05g/l _{culture}
(NH ₄) ₂ SO ₄	0.5g/l _{culture}

Oligo-éléments

FeCl ₃	207.5g/l _{culture}
ZnCl ₂	21g/l _{culture}
CuCl ₂	3.25g/l _{culture}
CoCl ₂	2.5g/l _{culture}
H ₃ BO ₃	2.5g/l _{culture}
EDTA	0.25g/l _{culture}

Source de carbone

Glucose	0.2% (w/v)
---------	------------

Vitamines

Biotine	1mg/ml _{culture}
Thiamine	1mg/ml _{culture}

Acides aminés

Leucine	200mg/l _{culture}
Isoleucine	200mg/l _{culture}
Valine	200mg/l _{culture}
Tryptophane	200mg/l _{culture}
Phénylalanine	200mg/l _{culture}
Lysine	200mg/l _{culture}
Alanine	200mg/l _{culture}
Arginine	200mg/l _{culture}
Acide aspartique	200mg/l _{culture}
Glycine	200mg/l _{culture}
Asparagine	200mg/l _{culture}
Proline	200mg/l _{culture}
Serine	200mg/l _{culture}
Thréonine	200mg/l _{culture}
Histidine	200mg/l _{culture}
Glutamine	200mg/l _{culture}
Acide glutamique	200mg/l _{culture}
Sélénométhionine	125mg/l _{culture}

ANNEXE 3: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DES DOMAINES 485-769 ET 503-769 DE SCMLH1

Production du CTD de scMlh1

Procédé

- **Transformation** : 2µL plasmide pETM30 MLH1 (f) clone 4 (cc 19/3/3) 1.23ng/µl (boîte LF1 plasmide réparation à -20°C) + 100µL bactéries compétentes BL21 Star DE3

30 min dans la glace

30 s à 42°C (eppendorf) ou 45s dans falcon

2-3 min dans la glace

Ajout 250µL de SOC

Incubation 250 rpm 45 min à 37°C

Ensemencement 100µl sur boîte de pétri LB + kanamycine (30mg/ml)

Incubation 37°C overnight (vérification si obtention de clones)

En parallèle mettre 200µl de la transformation dans 3ml de LB + 3µl de kanamycine (30mg/ml)

Incubation 250 rpm 5h à 37°C

- **Préculture** : dans un erlen de 500mL, 50mL de LB stéril + 50µL de kanamycine à 30mg/ml + 3 ml de bactéries transformées

Agitation 200rpm 37°C overnight

- Prise DO de la préculture au 1/10
- Ajout dans culture 1ml de kanamycine à 30mg/ml, ensemencement de 1L de **culture** avec de la préculture pour obtenir une DO=0,05

Incubation 37°C 200rpm

- **Induction** à DO= 1.2 avec IPTG à 0,5mM

Incubation 20°C 200rpm overnight

- Cultures transvasées dans JA 8.1000
- Centrifugation 4500rpm 40min 4°C
- Resuspendre les culots dans 30ml PBS 1X à 4°C
- Transvaser dans falcons 50ml
- Centrifugation 5000rpm 20min 4°C
- Culots conservés à -80°C

Purification du CTD de scMlh1

Matériel

⇒ Partir de culots de 2L congelés à -80°C

Tampons

Tp Lyse 250ml 50mM Na₂HPO₄ pH8
 500mM NaCl
 5% Glycérol
 1% Triton X-100
 1mM EDTA
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp P500 1L 50mM Na₂HPO₄ pH8
0,22μM 500mM NaCl
 1mM EDTA
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp élution GST tamponné pH8
P500 50mM Na₂HPO₄ pH8
0,22μM 500mM NaCl
 1mM EDTA
 10mM glutathion réduit
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp élution HisTrap tamponné pH8
P500 50mM Na₂HPO₄ pH8
0,22μM 500mM NaCl
 1mM EDTA
 0.5M imidazole
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp P300 1L 50mM Na₂HPO₄ pH8
0,22μM 300mM NaCl
 1mM EDTA
 10mM β-mercaptoéthanol
 0.5mM PMSF

Solutions stock:

6M Gnd-HCl
0.5M EDTA
Glycérol
0.1M NiSO₄

Colonnes

GSTrap HP 5ml
HiPrep Sephacryl S-200 26/60 B
HisTrap HP 5ml

PROCEDE

Lyse cellulaire

- ⇒ Le culot de 1L de culture est resuspendu dans 30ml de Tp de lyse.
- ⇒ Ajout de 25mg de lysozyme par litre de culture, laisser incuber 30min à 4°C sous agitation
- ⇒ Les cellules sont lysées par sonication (timer 1min, température:10°C, pulse: 1sec ON et 1 sec OFF, amplitude de 25% à 55% (écart de 10%))
- ⇒ Addition de 10mM Mg₂Cl et 1,5µl de benzonase (250U)
- ⇒ Incubation à 4°C pendant 15min
- ⇒ Centrifugation à 18000g JA20 pendant 30min
- ⇒ Filtration du surnageant à 0.45µm

Colonne d'affinité : GSTrap

Lavage de la colonne

- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 2CV de 6M Gnd-HCl à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min

Equilibration de la colonne

- ⇒ 10CV de Tp P500 à 1,5ml/min

Chargement de la colonne (60ml)

- ⇒ A un débit de 2ml/min, flow through collecté puis repassé sur la colonne

Lavage

- ⇒ Avec Tp P500 jusqu'à retrouver la ligne basal en mAU à 280. Collecter le lavage

Elution

- ⇒ 5CV de Tp P500Elution GSTrap.
- ⇒ Collection de fractions élués par fractions de 1ml (20ml pour première élution et 16ml pour l'élution du FT)
- ⇒ Quantification de la protéine par Bradford.

Coupure avec la TEV

- ⇒ Addition de 1% de TEV (protéine diluée afin d'éviter la précipitation) et incubation à 4°C sur la nuit avec agitation douce

Colonne d'affinité : HisTrap

Lavage et régénération de la colonne

- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 5CV de 0.5M EDTA à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 3ml de 0.1M NiSO₄ à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 2.5ml de 0.1M NiSO₄ à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min

Equilibration de la colonne

- ⇒ 5CV de Tp P500

Chargement de la colonne et élution de la protéine

- ⇒ A un débit de 2.5ml/min
- ⇒ Collection du flow through
- ⇒ Remettre dans Tp P500

Elution des protéines avec étiquette his

- ⇒ 5CV de Tp P500 Elu HisTrap, élution collectée

- Quantification du FT récolté par nanodrop

- Concentration des fractions des protéines

- ⇒ Concentration par cellule amicon, Memb 3 000Da
- ⇒ Vérification de la quantité de protéine par Nanodrop au cours de la concentration

Gel filtration: HiPrep Sephacryl S-200 26/60 B

Equilibration de la colonne

- ⇒ 1CV de Tp P300 en présence d'EDTA à 0.6 ml/min

Chargement de la colonne et élution de la protéine

- ⇒ Injection de 10ml maximum sur la colonne HiPrep Sephacryl S-200 26/60 B à 0,6ml/min
- ⇒ Collection des fractions par 2ml à 0,6ml/min

Quantification de la protéine

- ⇒ Nous avons mesuré la concentration du pool collecté par Nanodrop

ANNEXE 4: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DU DOMAINE 486-756 DE hSMlh1

Production du CTD de hSMlh1

Procédé

- **Transformation** : 2µL plasmide H5-7 pETM30 clone 5 (MLH1 Humain 24/05/2005 150ng/µl) (boîte LF1 plasmide réparation à -20°C) + 100µL bactéries compétentes BL21 Star DE3

30 min dans la glace

30 s à 42°C (eppendorf) ou 45s dans falcon

2-3 min dans la glace

Ajout 250µL de SOC

Incubation 250 rpm 45 min à 37°C

Ensemencement 100µl sur boîte de pétri LB + kanamycine (30mg/ml)

Incubation 37°C overnight (vérification si obtention de clones)

En parallèle mettre 200µl de la transformation dans 3ml de LB + 3µl de kanamycine (30mg/ml)

Incubation 250 rpm 5h à 37°C

- **Préculture** : dans un erlen de 500mL, 50mL de LB stéril + 50µL de kanamycine à 30mg/ml + 3 ml de bactéries transformées

Agitation 200rpm 37°C overnight

- Prise DO de la préculture au 1/10
- Ajout dans culture 1ml de kanamycine à 30mg/ml, ensemencement de 1L de culture avec de la préculture pour obtenir une DO=0,05

Incubation 37°C 200rpm

- **Induction** à DO= 1.2 avec IPTG à 0,5mM

Incubation 20°C 200rpm overnight

- Cultures transvasées dans JA 8.1000
- Centrifugation 3500rpm 30min 4°C
- Resuspendre les culots dans 30ml PBS 1X à 4°C
- Transvaser dans falcons 50ml
- Centrifugation 5000rpm 20min 4°C
- Culots conservés à -80°C
-

Purification du CTD de hsm1h1

Matériel

⇒ Partir de culot de 1L congelé à -80°C

Tampons

Tp Lyse 250ml 50mM Na₂HPO₄ pH8
 500mM NaCl
 5% Glycérol
 1% Triton X-100
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp P500 1L 50mM Na₂HPO₄pH8
0,22μM 500mM NaCl
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp élution GST tamponné pH8
P500 50mM Na₂HPO₄ pH8
0,22μM 500mM NaCl
 10mM glutathion réduit
 Attention remettre sous argon ou azote la poudre de glutathion
 après pesée
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp élution HisTrap tamponné pH8
P500 50mM Na₂HPO₄ pH8
0,22μM 500mM NaCl
 0.5M imidazole
 10mM β-mercaptoéthanol
 0.5mM PMSF

Tp P300 1L 50mM Na₂HPO₄ pH8
0,22μM 300mM NaCl
 1mM EDTA
 10mM β-mercaptoéthanol
 0.5mM PMSF

Solutions stock:

6M Gnd-HCl

Glycérol
0.1M NiSO₄

Colonnes

GSTrap HP 5ml
HiPrep Sephacryl S-200 26/60 B
HisTrap HP 5ml

Procédé

Lyse cellulaire

- ⇒ Le culot de 1L de culture est resuspendu dans 30ml de Tp de lyse.
- ⇒ Ajout de 25mg de lysozyme (PM 14.3kDa) par litre de culture, laisser incuber 30min à 4°C sous agitation
- ⇒ Les cellules sont lysées par sonication (timer 1min, température:10°C, pulse: 1sec ON et 1 sec OFF, amplitude de 25% à 55% (écart de 10%))
- ⇒ Addition de 10mM Mg₂Cl et 1,5µl de benzonase (250U/ul, 375U)
- ⇒ Incubation à 4°C pendant 15min
- ⇒ Centrifugation à 18000g JA20 pendant 30min
- ⇒ Filtration du surnageant à 0.45µm

Colonne d'affinité : GSTrap vol=5ml

Lavage de la colonne

- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 2CV de 6M Gnd-HCl à 2ml/min, vérifier la valeur du pH
- ⇒ 5CV de H₂O à 5ml/min

Equilibration de la colonne

- ⇒ 10CV de Tp P500 à 1,5ml/min

Chargement de la colonne

- ⇒ A un débit de 2ml/min, flow through collecté puis repassé sur la colonne plus tard

Lavage

- ⇒ Avec Tp P500 jusqu'à retrouver la ligne basal en mAU à 280. Collecter le lavage

Elution

- ⇒ 5CV de Tp P500Elution GSTrap.
- ⇒ Collection de fractions élués par fractions de 1ml
- ⇒ Quantification de la protéine par Bradford.

Coupure avec la TEV

- ⇒ Ramener la concentration en fusion Gst-Mlh1 à 0,5-1mg/ml
- ⇒ Addition de 1% de TEV (protéine diluée afin d'éviter la précipitation) et incubation à 4°C overnight avec agitation douce.

Colonne d'affinité : HisTrap

Lavage et régénération de la colonne

- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 5CV de 0.5M EDTA à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 3ml de 0.1M NiSO₄ à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min
- ⇒ 2.5ml de 0.1M NiSO₄ à 2ml/min
- ⇒ 5CV de H₂O à 5ml/min

Equilibration de la colonne

- ⇒ 5CV de Tp P500

Chargement de la colonne et élution de la protéine

- ⇒ A un débit de 2.5ml/min
- ⇒ Collection du flow through
- ⇒ Remettre dans Tp P500

Elution des protéines avec étiquette his

- ⇒ 5CV de Tp P500 Elu HisTrap, élution collectée

- Quantification du FT récolté par nanodrop (ou Bradford si spectre est non exploitable pas d'épaulement à 280nm)

- Concentration des fractions des protéines

- ⇒ Concentration par cellule amicon, Memb 3 000Da
- ⇒ Vérification de la quantité de protéine par Nanodrop au cours de la concentration

Gel filtration: HiPrep Sephacryl S-200 26/60 B

Equilibration de la colonne

- ⇒ 1CV de Tp P300 en présence d'EDTA à 0.6 ml/min

Chargement de la colonne et élution de la protéine

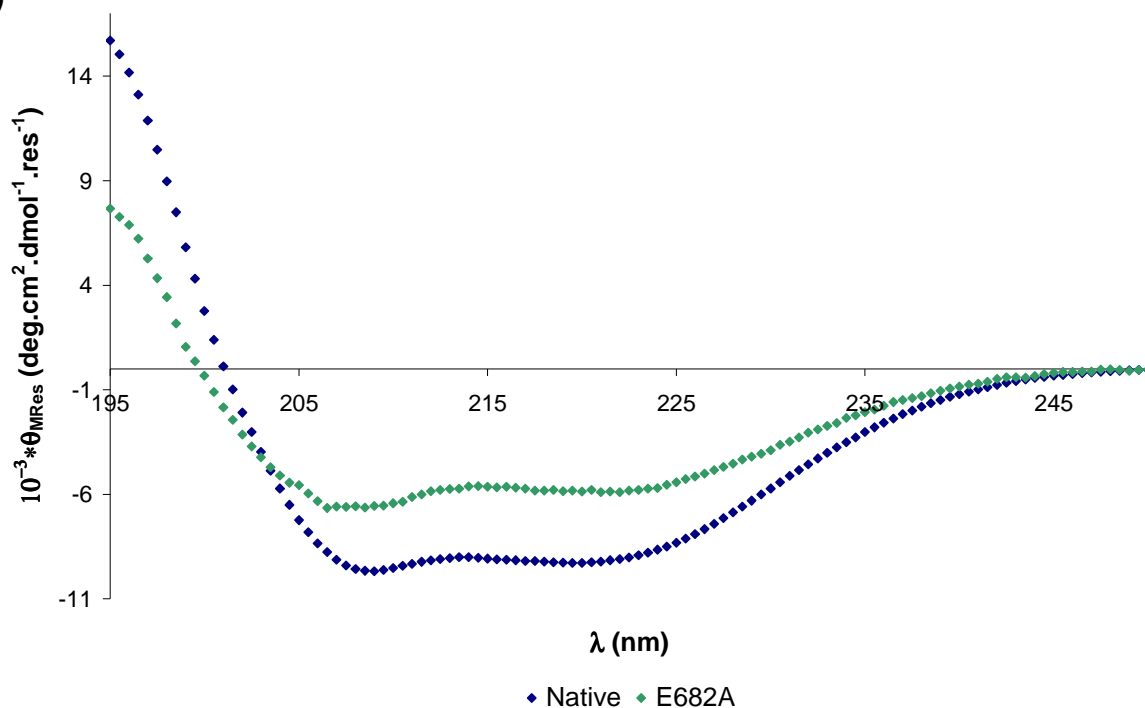
- ⇒ Injection de 10ml maximum sur la colonne HiPrep Sephacryl S-200 26/60 B à 0,6ml/min
- ⇒ Collection des fractions par 2ml à 0,6ml/min

Quantification de la protéine

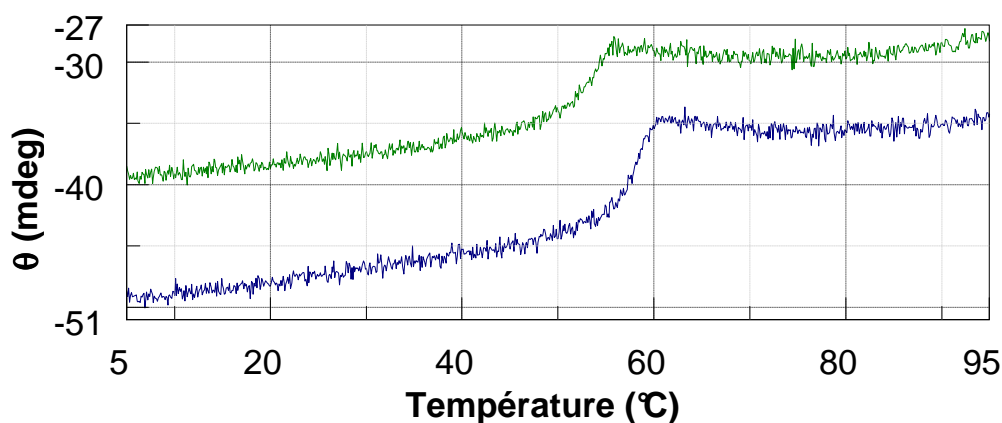
- ⇒ Nous avons mesuré la concentration du pool collecté par **Nanodrop**

ANNEXE 5: SPECTRE DE DICHROISME CIRCULAIRE ET DENATURATION THERMIQUE DU VARIANT E682A DU DOMAINE C-TERMINAL DE MLH1

(A)

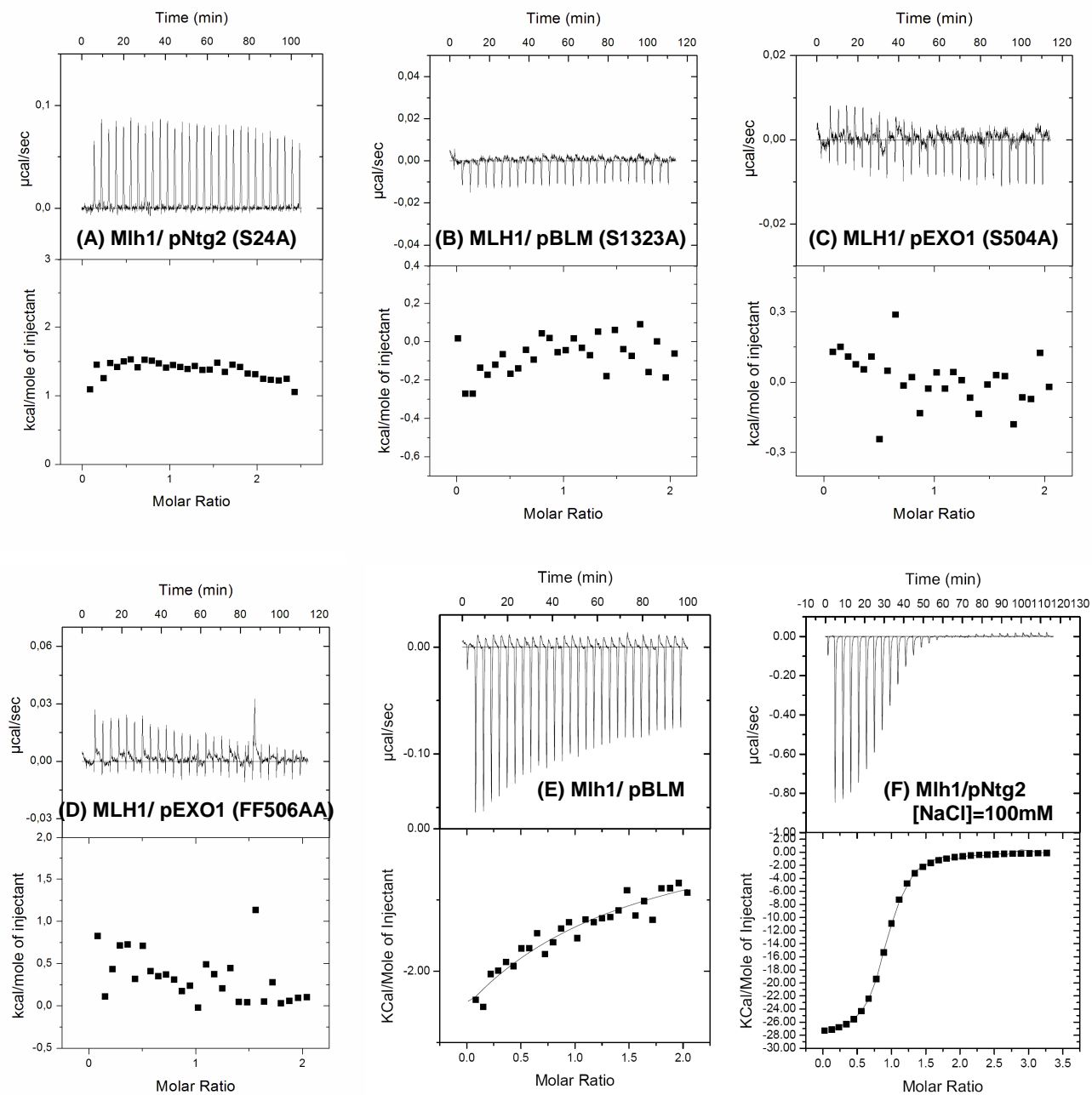


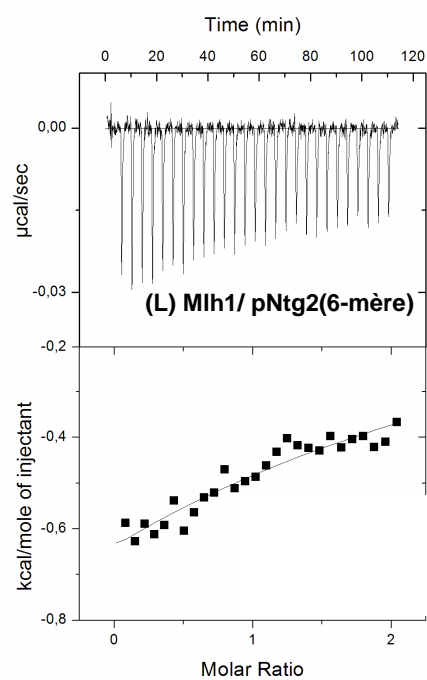
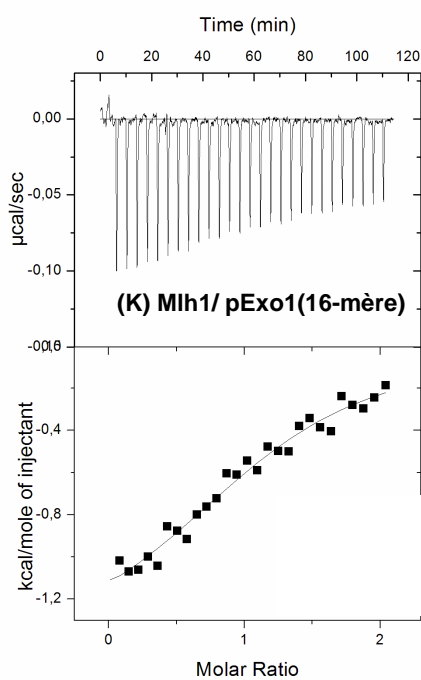
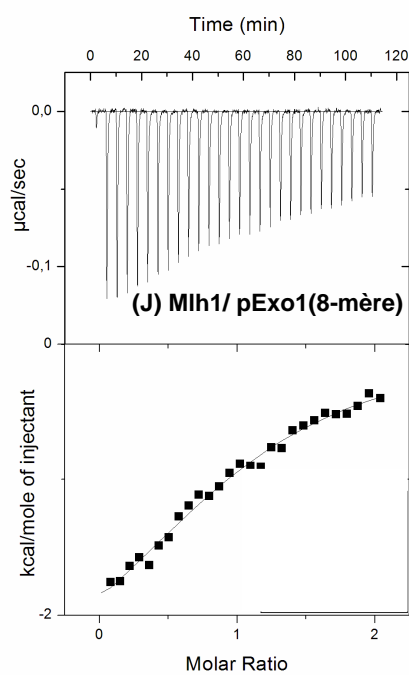
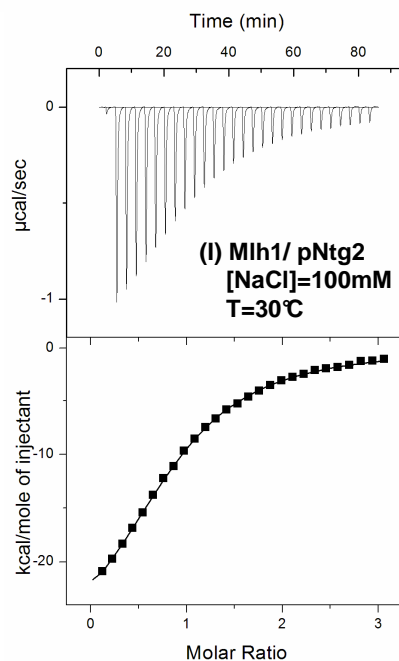
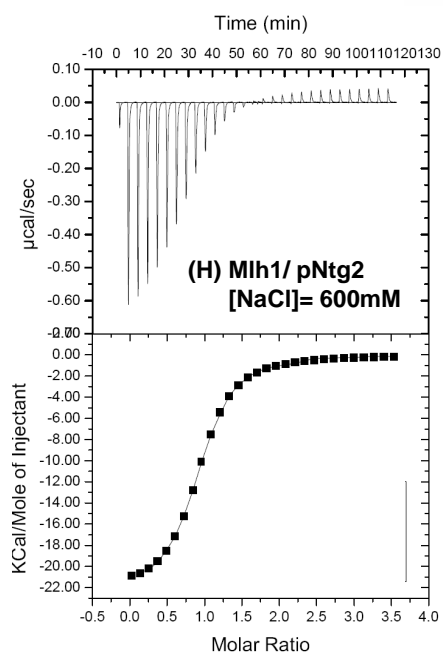
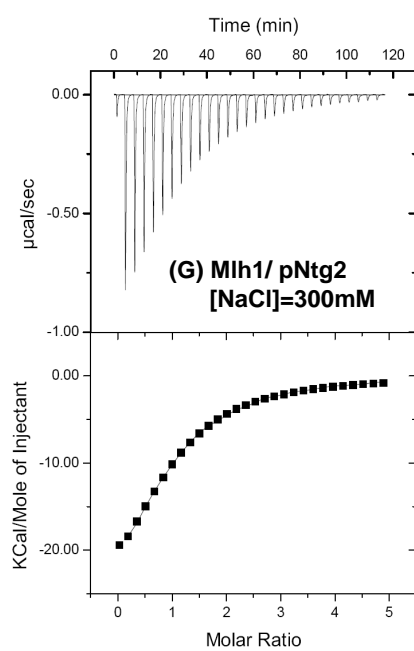
(B)

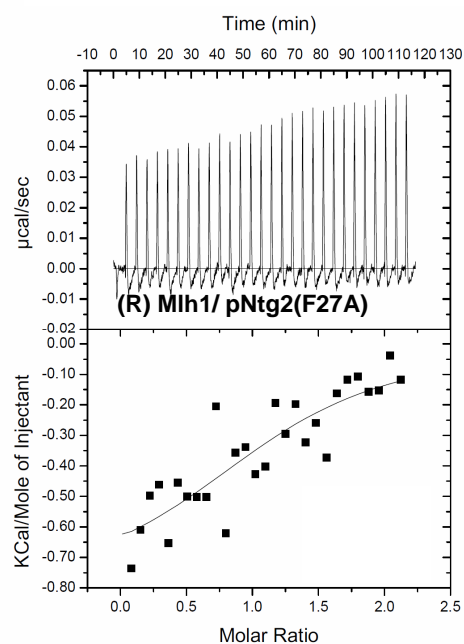
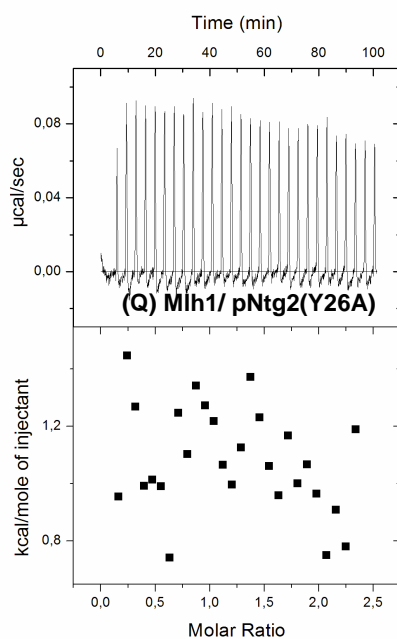
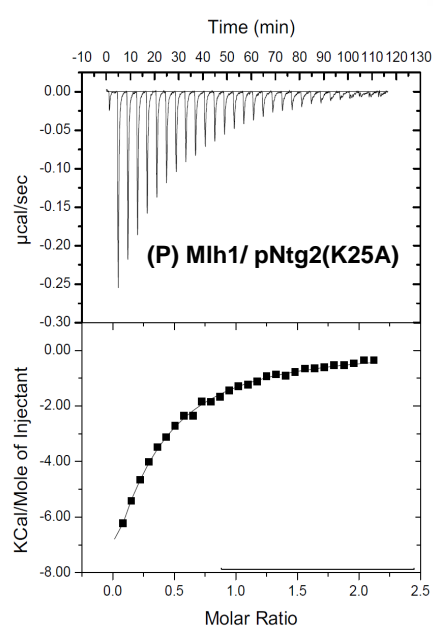
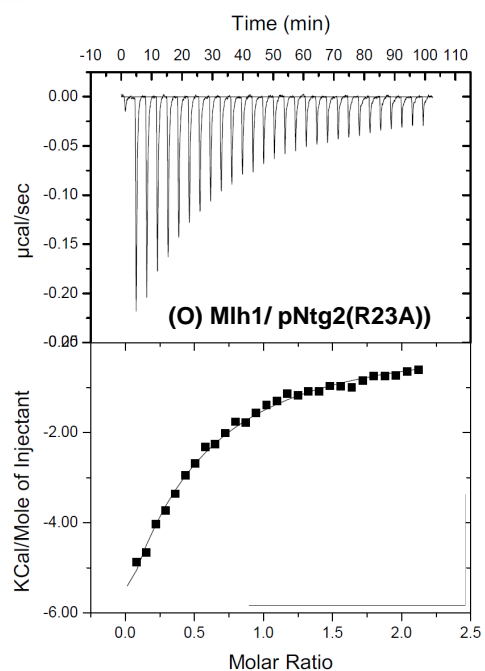
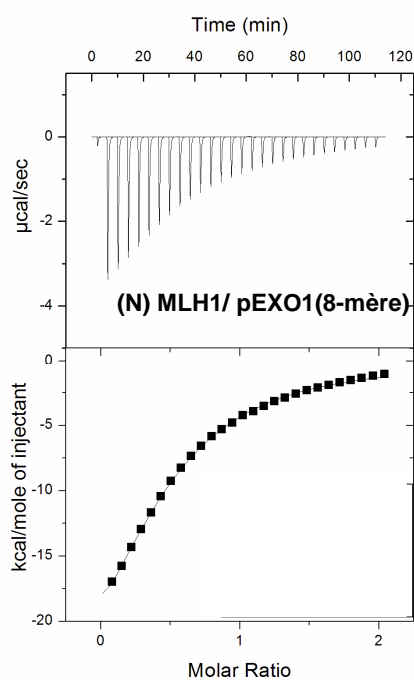
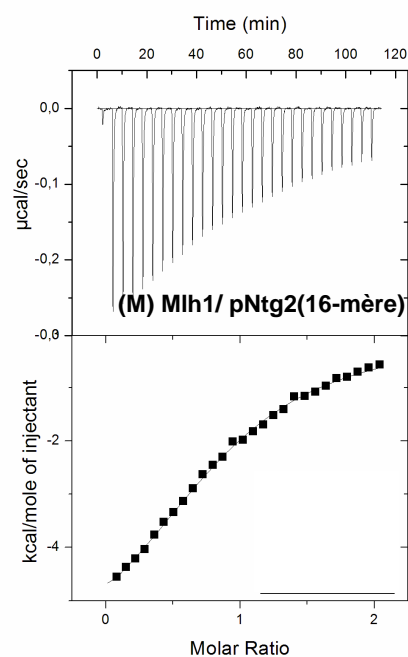


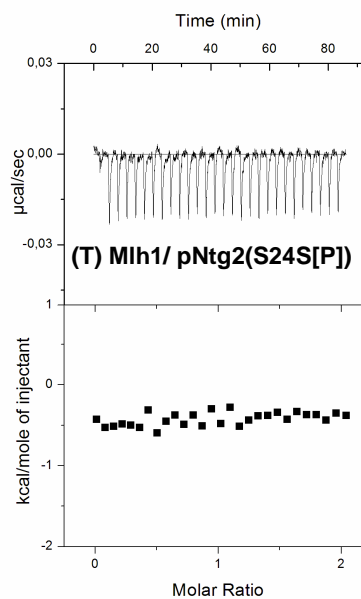
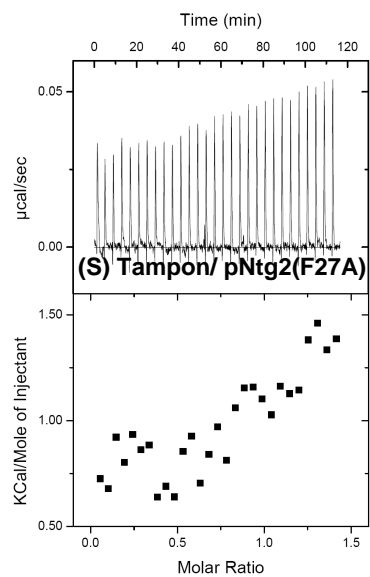
(A) Spectre de dichroïsme circulaire du CTD de Mlh1 native et du variant E682A. Les spectres de la protéine native (en bleu) et du mutant E682A (en vert) sont obtenus à partir d'échantillons concentrés tels que: [native]=7 μ M et [E682A]=3 μ M **(B)** Dénaturation thermique du CTD de Mlh1 native (en bleu) et du variant E682A (en vert)

ANNEXE 6: THERMOGRAMME ET ISOTHERME DU DOMAINE C-TERMINAL DE MLH1 ET DE PEPTIDES DERIVES DU MOTIF MIP

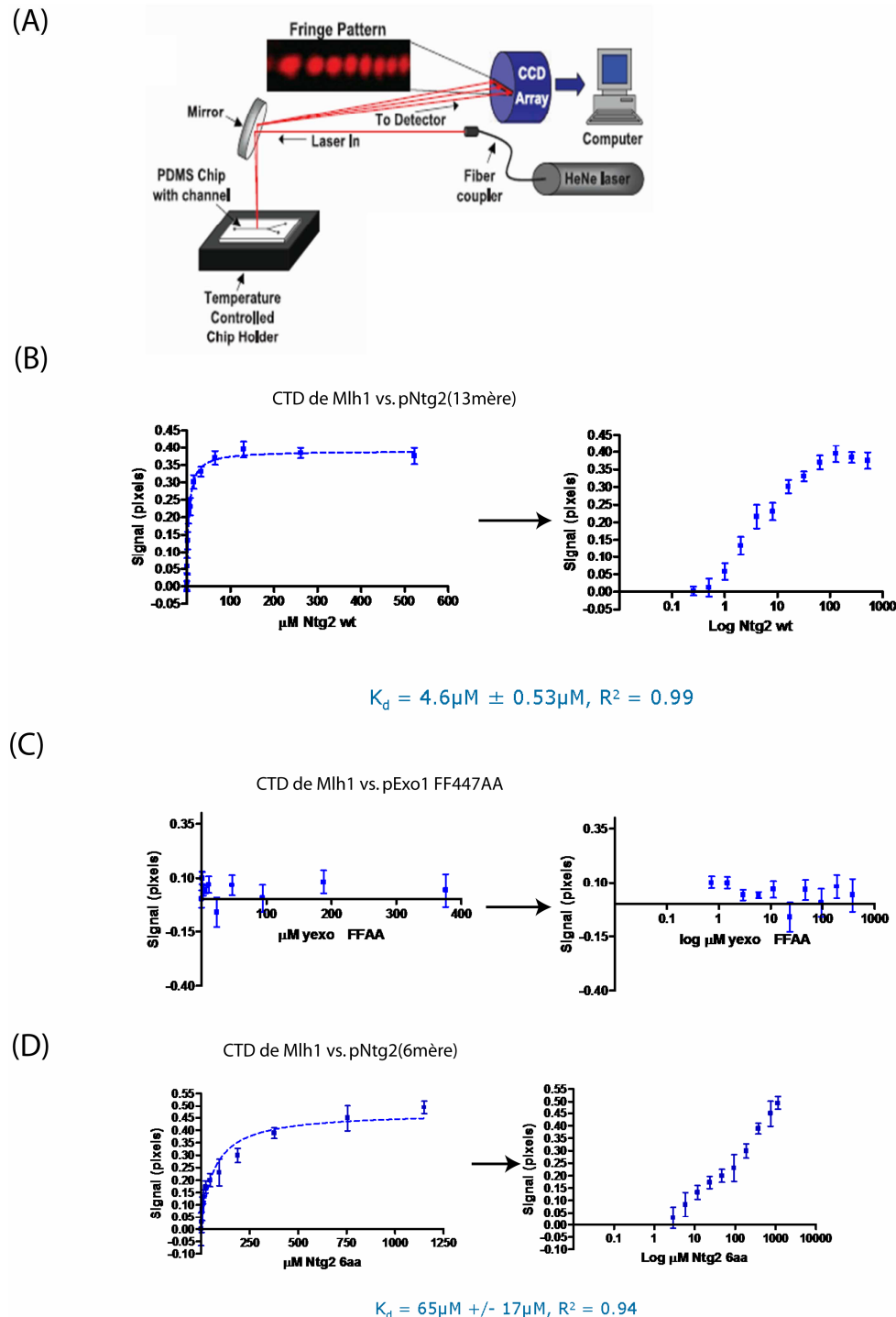








ANNEXE 7: CARACTERISATION PAR BSI DES INTERACTIONS DE PEPTIDES CONTENANT UN MOTIF MIP AVEC LA REGION C-TERMINALE DE MLH1



Mesure d'interaction par BSI (A) Schéma du montage expérimental de la BSI (Back Scattering Interferometry) (d'après (Bornhop, 2007 Science)) (B) Courbe expérimentale d'interaction entre le CTD de Mlh1 à 10μM et le peptide pNtg2(13mère) à 100μM (C) Courbe expérimentale d'interaction entre le CTD de Mlh1 à 10μM et le peptide pExo1-FF447AA à 100μM

ANNEXE 8: CONDITIONS DE CRISTALLOGENESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1 (485-769)

Type de Boîte	Condition	Réagent numér	Sel	pH	Tampon	Agent précipitant	Agent précipitant 2	Observations	Goutte
W1	C			7	0.01 M Bis-Tris	20% (w/v) PEG 4000			
Q3	a08	98	0.1 M Sodium sulfate			30% (w/v) PEG 3350		Pts ou précipité homogène	
Q2	a08	99	0.2 M Lithium sulfate	6,5	0.1 M TRIS-HCl	30% (w/v) PEG 4000		Sph ou SP (+)	Soluble
Q2	a07	99	0.2 M Ammonium sulfate			30% (w/v) PEG 4000		Pts hétéro ++, sph?	
Q1	a11	99		7,5	0.1 M HEPES sodium salt	1.8 M Lithium sulfate		Sph ou SP (+)	
Q3	a01	9		4,0	0.1 M Sodium acetate	1.8 M Ammonium sulfate		Sph	
Q1	a09	9	0.2 M tri-Sodium citrate	6,5	0.1 M Sodium cacodylate	30% (v/v) Isopropanol		Pts hétéro ++; flou?	
Q1	a08	32	0.1 M Sodium chloride	7,5	0.1 M HEPES	1.8 M Ammonium sulfate		Sph ou pts hétéro ++; flou?	
Q2	a12	93	0.2 M tri-Lithium citrate			20% (w/v) PEG 3350		Sph ou SP?	Pts léger
Q3	a07	49	0.01 M Magnesium chloride	5,6	0.05 M MES	2 M Lithium sulfate		4 sph	L soluble, R: pts
Q1	a08	8	0.2 M Calcium chloride	4,8	0.1 M Sodium acetate	20% (v/v) Isopropanol		Pts halo	Pts
Q1	a10	10	0.2 M Magnesium chloride	7,5	0.1 M HEPES sodium salt	30% (v/v) Isopropanol		Cristaux irréguliers; sel?	Pts
Q1	a07	79	0.01 M Zinc sulfate	6,5	0.1 M MES	25% (w/v) PEG 650 MME		porétal	Léger pts
Q2	b10	74	0.2 M Zinc acetate			20% (w/v) PEG 3350		Pts hétéro; amas	Pts
Q2	a04	31	0.2 M Ammonium sulfate			15% (w/v) PEG 4000		Pts hétéro; amas	Pts
Q2	a09	72	0.2 M Lithium acetate			20% (w/v) PEG 3350		Sph	Léger pts
Q3	b01	2	0.1 M Zinc acetate	4,8	0.1 M Sodium acetate	12% (w/v) PEG 4000		SP ou précipité; sels?	
Q3	a12	92	0.2 M Potassium chloride	7,5	0.05 M TRIS-HCl	0.05 M Magnesium chloride	10% (w/v) PEG 4000	cristaux; sels?	Soluble
Q3	a10	77	0.01 M Magnesium chloride	7,0	0.05 M HEPES Sodium salt		4 M Lithium chloride	Sph; sels?	Soluble
Q3	a11	88	0.2 M Ammonium acetate	7,0	0.05 M HEPES Sodium salt	0.15 M Magnesium acetate	5% (w/v) PEG 4000	Cristaux ou pous; sels?	Soluble
Q3	a12	94	0.005 M Magnesium sulfate eq.	6,5	0.05 M TRIS-HCl		35% (w/v) 1,6-Hexanediol	Cristaux au bord; sels?	Soluble
Q3	a04	31	0.2 M Magnesium chloride	7,5	0.1 M HEPES sodium salt	18% (w/v) PEG 400		Sp ou cristaux; sels?	
P4	a08	50	0.8 M Na/K phosphate	5,8					True
P4	a03	21	3.2 M Ammonium sulfate	6,0	0.1 M MES			Sph	Léger pts
P4	a05	30	3.4 M Sodium malonate	4,0				Sph	Soluble
Q5	a11	77	1.0 M Lithium chloride	8,0	0.1 M TRIS			Sph	Soluble
Q5	a12	78	1.0 M Lithium chloride	9,0	0.1 M BICINE			Sph	Soluble
Q5	b12	84	1.0 M Lithium chloride	9,0	0.1 M BICINE	10% (w/v) PEG 8000		Sph	Soluble
Q5	a08	88	1.0 M Lithium chloride	5,0	0.1 M Citric acid	20% (w/v) PEG 8000		Sph ou pts hétéro ++; flou?	Pts
Q5	a10	88	1.0 M Lithium chloride	7,0	0.1 M HEPES	20% (w/v) PEG 8000		Sph	Soluble
Q5	a11	89	1.0 M Lithium chloride	8,0	0.1 M TRIS	20% (w/v) PEG 8000		Sph	Soluble
Q5	a12	90	1.0 M Lithium chloride	9,0	0.1 M BICINE	20% (w/v) PEG 8000		Sph ou précipité	Soluble
Q5	a08	92	1.0 M Lithium chloride	5,0	0.1 M Citric acid	30% (w/v) PEG 8000		Pts hétéro; amas	Pts
Q5	a09	93	1.0 M Lithium chloride	6,0	0.1 M MES	20% (w/v) PEG 8000		Sph	Pts
Q5	a10	94	1.0 M Lithium chloride	7,0	0.1 M HEPES	30% (w/v) PEG 8000		Sph	Pts
Q5	a11	95	1.0 M Lithium chloride	8,0	0.1 M TRIS	30% (w/v) PEG 8000		Sph	Pts
Q5	a12	96	1.0 M Lithium chloride	9,0	0.1 M BICINE	30% (w/v) PEG 8000		Sph ou précipité	Pts
Q5	a08	54		9,0	0.1 M BICINE	10% (v/v) MPD		Sph	Soluble
Q5	a07	19		4,0	0.1 M Citric Acid	20% PEG 5000 Mma		Pts hétéro; amas	Pts
Q5	a08	20		5,0	0.1 M Citric Acid	20% PEG 5000 Mma		Pts hétéro; amas	Pts
Q6	a04	26	3.6 M Sodium formate	7,0				1 SPH	Soluble
Q6	a07	49	0.2 M Calcium chloride	6,5	0.1 M BIS-TRIS	45% (w/v) MPD		1 cristal en losange?	Léger pts
Q6	b08	42	0.1 M BIS-TRIS	5,5	25% (w/v) PEG 3350			Sp ou cristaux; sels?	Léger pts?
Q6	a10	78	0.2 M Lithium sulfate	7,5	0.1 M HEPES	25% (w/v) PEG 3350		4 µsph ?	
Q6	a10	77	0.2 M Lithium sulfate	6,5	0.1 M TRIS	25% (w/v) PEG 3350		sph	Soluble
Q6	a08	63	0.064 M Sodium citrate	7,0	0.1 M HEPES	10% (w/v) PEG 5000 MME		1 porétal	Soluble

En bleu = Conditions ayant données des microcristaux, des sphérulites reproduites au LBSR.

En vert = Conditions ayant données des sphérulites ou du précipité hétérogène de qualité inférieur et non reproduites au LBSR.

Les autres conditions ont données des résultats intéressants mais soit de moins bonne qualité que les conditions bleues ou vertes, soit sujettes au doute sur la nature des cristaux ou du sphérulites.

ANNEXE 9: CONDITIONS DE CRISTALLOGENESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1 (503-769)

ecre en ram e	Well	Reage nt num er	salt	buffer	pH	Agent précipitant	Agent précipitant 2	Obs	no te
JCS G	A10	1,84	0.2 M Potassium formate			20 % (w/v) PEG 3350		L: pté hetero; M: lg pté; R: ptées aiguilles, app à j+3	4
JCS G	B4	1,04		0.1 M HEPES	7,5	10 % (w/v) PEG 8000	8 % (v/v) Ethylene glycol	L: sep de q; M: aiguille, app à j+6; R: sol	4
JCS G	C7	2,81	0.2 M Zinc acetate	0.1 M Sodium acetate	4,5	10 % (w/v) PEG 3000		L, M, R: sph+pté homo, no evol	4
JCS G	E6	3,46	0.2 M Zinc acetate	0.1 M Imidazole	8,0	20 % (w/v) PEG 3000		L, M, R: sph+pté homo, no evol	4
JCS C	C8	4,82	0.15 M DL-Malic acid pH 7.0			20 % (w/v) PEG 3350		L: aiguilles, no evol; M, R: aiguilles, app à j+3	4
Q2	a03	17	0.2 M Lithium sulfate	0.1 M TRIS.HCl	8,5	15 % (w/v) PEG 4000		L: aiguilles, app à j+3; M: aiguilles, app à j+6; R: sol	4
Q2	f08	62	0.2 M Potassium thiocyanate			20 % (w/v) PEG 3350		L, M: sep de q; R: aiguille, app à j+8	4
Q2	f03	22	0.2 M Sodium acetate	0.1 M TRIS.HCl	8,5	15 % (w/v) PEG 4000		L: aiguille à 1 mois; M: xto ou pté het, no evol; R: sol	4
Q2	h06	40		0.1 M tri-Sodium citrate	5,6	10 % (v/v) Isopropanol, 10 % (w/v) PEG 4000		L: M: aiguille à j+10; R: sol	4
Q3	b05	34	0.1 M Ammonium sulfate	0.1 M HEPES sodium salt	7,5	10 % (w/v) PEG 4000		L, M: aiguilles, app à j+3; R: sol	4
Q3	d06	44	0.1 M di-Ammonium phosphate	0.1 M TRIS	8,5	12 % (w/v) PEG 6000		L: M: aiguilles, app à j+3; R: sol	4
Q3	e07	53	0.2 M Potassium chloride	0.05 M MES	5,6	0.01 M Magnesium chloride	5 % (w/v) PEG 8000	L: aiguilles, app à j+8; M: xto à j+20; R: sol	4
Q3	f04	30	0.1 M Sodium chloride	0.1 M HEPES sodium salt	7,5	10 % (w/v) PEG 4000		L: aiguilles, app à j+3; M: aiguilles ou pté; R: sol	4
Q3	f11	86	0.2 M Ammonium acetate	0.05 M HEPES Sodium salt	7,0	0.15 M Magnesium acetate	5 % (w/v) PEG 4000	L: plaquette, app à j+3; M: id à j+8; R: sol	4
Q3	a12	88	0.1 M Potassium chloride	0.05 M TRIS.HCl	7,5	0.015 M Magnesium chloride	10 % (w/v) PEG MMC 350	L, M, R: ptit aiguille, app à j+3	4
Q3	a06	57	0.1 M Potassium chloride	0.05 M MES	6,0	0.01 M Magnesium chloride	10 % (v/v) PEG 400	L, M, R: aiguille à j+20	4
Q3	b03	18	0.1 M Lithium sulfate	0.1 M tri-Sodium citrate	5,6	12 % (w/v) PEG 6000		L: sep de q; M: sep de q; R: aiguille à j+20	4
Q3	b10	74	0.2 M Potassium chloride	0.05 M Sodium cacodylate	6,5	0.1 M Magnesium acetate	10 % (w/v) PEG 8000	L: pté? ds sep de q; M: R: ptit aiguille à j+20	4
Q3	d02	12	0.1 M Sodium chloride	0.1 M tri-Sodium citrate	5,6	18 % (v/v) PEG 400		L: Xto 'galets' à 1 mois; M, R: sol	4
Q3	d07	52	0.2 M Potassium chloride	0.05 M MES	5,6	0.01 M Magnesium sulfate	10 % (v/v) PEG 400	L: plaquette (macie?) à 1 mois; M, R: sol	4
Q3	g03	23	0.1 M Lithium sulfate	0.1 M ADA	6,5	12 % (w/v) PEG 4000, 2 % (v/v) Isopropanol		L, M, R: aiguille à j+20	4
Q3	g04	31	0.1 M Magnesium chloride	0.1 M HEPES sodium salt	7,5	18 % (v/v) PEG 400		L: xto ou pté het?; M, R: tapis ptit aiguille à j+20	4
Q6	e12	6		0.1 M Bisine	8,0	5% PEG 6000 Mme		L: plaquettes à j+8; M, R: sol	4
Q6	e06	41		0.1 M TRIS	8,0	20 % (w/v) PEG 6000		L: pté; M, R: aiguilles, no evol	4
Q6	a12	86	0.1 M Succinic acid pH 7.0			15 % (w/v) PEG 3350		L: ptite aiguille; evol jusqu'à j+8; M: sol; R: ptit aiguille à 1 mois	4
Q6	e12	91	0.15 M DL-Malic acid pH 7.0			20 % (w/v) PEG 3350		L: sep de q; M, R: aiguilles, evol sur R à j+3	4
Q6	d12	92	0.1 M Magnesium formate			15 % (w/v) PEG 3350		L, M: sep de q; R: ptées aiguilles, no evol	4
Q6	e08	61	0.2 M L-Proline	0.1 M HEPES	7,5	10 % (w/v) PEG 3350		L: aiguilles, evol j+3; M: aiguilles, evol j+8; R: aiguille à 1 mois	4
Q6	g08	83	0.064 M Sodium citrate pH 7.0	0.1 M HEPES	7,0	10 % (w/v) PEG 5000 MME		L: M: aiguilles, app à j+6; R: sol	4
Q6	b12	90	0.2 M Sodium formate			20 % (w/v) PEG 3350		L: M: pté+sep de f; R: aiguille, app à j+3	4

JCS G	A1	1,A1	0.2 M Lithium sulfate	0.1 M Sodium acetate	4,5	50 %(v/v) PEG 400	L, M: pté homo; R:xtal ou fibro+pté homo, no evol	3
JCS G	A2	1,A2		0.1 M tri-Sodium citrate	5,5	20 %(w/v) PEG 3000	L,M:sol, R:uxto ou pté translu,evol à j+3	3
JCS G	A12	1,B6	0.2 M Potassium nitrate			20 %(w/v) PEG 3350	L:1 gros amas?+sol, no evol; M: sol; R: lg pté het	3
JCS G	B12	1,D6	0.2 M tri-Potassium citrate			20 %(w/v) PEG 3350	L: M: pté hetero; R: aiguille ou lg pté hetero, app à j+6	3
JCS G	C12	2,B6				10 %(w/v) PEG 1000 10% (w/v) PEG 8000	L,M:pté het+homo; R: 1 xtal?+pté het+homo; no evol	3
JCS G	D3	2,C3	0.2 M Sodium chloride	0.1 M Na/K phosphate	6,2	50 %(v/v) PEG 200	L: uxto ou pté translu, app à j+3; M,R:sol	3
JCS G	G1	4,A1		0.1 M HEPES	7,0	30 %(v/v) Jeffamine ED-2001	L,R:sol; M: lg aiguilles ou peau, app à j+6	3
JCS G	G4	4,A4	0.2 M Trimethylamine N-oxide	0.1 M TRIS	8,5	20 %(w/v) PEG MME 2000	L: pté hetero; M:uxto ou pté translu, app à j+3; R:sol	3
JCS G	A3	1,A3	0.18 M tri-Ammonium citrate			20 %(w/v) PEG 3350	L:uxto ou pté translu, M,R:sol	3
JCS G	G2	4,A2	0.02 M Magnesium chloride	0.1 M HEPES	7,5	22 %(w/v) Polyacrylic acid 5100 sodium salt	L:uto ou pté translu? a 1 mois;M,R: sol	3
PA CT	A1	1,A1		0.1M SPG buffer	4,0	25% w/v PEG 1500	L,M:pté homo; R: obj non idl; no evol	3
PAC T	C11	2,B5	0.2M Calcium chloride	0.1M Hepes	7,0	20% w/v PEG 8000	L:sep de φ+uxto ou bulle?;M:sep de φ;R:sol; no evol	3
PAC T	E8	3,A6	0.2M Sodium formate			20% w/v PEG 3350	L:sep de φ+1 xtal ou bulle?;M,R:sol; no evol	3
PAC T	E12	3,B6	0.2M Sodium malonate			20% w/v PEG 3350	L,M:pté hetero; R:pté, amas; no evol	3
Q1	a01	1	0.01 M Cobalt chloride	0.1 M Sodium acetate	4,6	1.0 M 1,6-Hexanediol	L,M:pté hetero; R: pté hetero + xto ou amas	3
Q1	a05	5		0.1 M HEPES sodium salt	7,5	10 %(v/v) Isopropanol, 20 %(w/v) PEG 4000	L: rien! ; M: pté hetero+ amas; R:pté hetero+1 amas	3
Q1	c06	30		0.1 M TRIS.HCl	8,5	2.0 M Ammonium sulfate	L: rien! ; M: Lg pté hetero; R:uxto? + lg pté, no evol	3
Q1	d05	65	0.5 M Lithium sulfate			15 %(w/v) PEG 8000	L: pté hetero forte+1xtal?, no evol ; M, R:pté homo	3
Q2	a02	9	0.2 M Ammonium acetate	0.1 M tri-Sodium citrate	5,6	15 %(w/v) PEG 4000	L: sep de φ; M: Xtal?+sol; no evol;R:sol	3
Q2	b03	18	0.2 M Magnesium acetate	0.1 M Sodium cacodylate	6,5	10 %(w/v) PEG 8000	L,M: sep de φ; R:amas ou uxto?, app à j+3	3
Q2	b09	66	0.2 M Potassium nitrate			20 %(w/v) PEG 3350	L: sep de φ; M: pté ou uxto?, app à j+3; R:sol	3
Q2	b10	74	0.2 M Zinc acetate			20 %(w/v) PEG 3350	L: pté homo; M,R: sph ou bulle+lg pté homo, no evol	3
Q2	g07	55	0.2 M Calcium chloride			20 %(w/v) PEG 3350	L:xtal?+pté hetero, no evol ;M,R:pté hetero	3
Q2	h10	80	0.2 M Magnesium sulfate			20 %(w/v) PEG 3350	L,M,R: amas, no evol	3
Q3	d11	84	0.1 M Potassium chloride	0.05 M HEPES Sodium salt	7,0	0.01 M Calcium chloride 10 %(v/v) PEG 400	L,M:uxto ou pté het?; R:sol	3
Q4	c11	65	1.4 M Na/K phosphate		7,5		L: rien!; M:1 xtal+sol, no evol; R: sol	3
Q5	b10	82	1.0 M Lithium chloride	0.1 M HPPFS	7,0	10 %(w/v) PEG 8000	L:sep de φ;M:uxto à j+8; R:sol	3
Q5	e08	54		0.1 M BICINE	8,0	10 %(v/v) MPD	L:1 xtal ou bulle?, no evol; M,R:sol	3
Q5	h08	20		0.1 M Citric Acid	5,0	20% PEG5000 Mne	L: 1 pté hetero; M: amas, no evol; R: amas + faible	3

En rouge = Conditions ayant données des cristaux

En jaune = Conditions ayant données des microcristaux, des sphérulites ou du précipité cristallin

ANNEXE 10: PROTOCOLE DE PRODUCTION ET DE PURIFICATION DES DOMAINES 485-769 DE SCMLH1*635-873 DE SCPMS1

Production du CTD de scMutL α

Procédé

- **Transformation** : 2 μ L plasmide PM1 pExp PMS1 166ng/ μ L CQ 16.06.01 + 100 μ L bactéries BL21 Star DE3 compétentes pACYC177-Mlh1

30 min dans la glace

30 s à 42°C (eppendorf) ou 45s dans falcon

2-3 min dans la glace

Ajout 250 μ L de SOC

Incubation 250 rpm 45 min à 37°C

Ensemencement 100 μ L sur boîte de pétri LB + kanamycine (30mg/ml) + ampicilline (50mg/ml)

Incubation 37°C overnight (vérification si obtention de clones)

- **Préculture** : dans un erlen de 500mL, 100mL de LB stérile + 100 μ L de kanamycine à 30mg/ml + 100 μ L d'ampicilline à 50mg/ml + quelques clones

Agitation 200rpm 37°C fin de journée + overnight

- Prise DO de préculture au 1/10
- Ajout dans culture 1ml de kanamycine et 1ml d'ampicilline, ensemencement des culture de 1L LB + 2g/L de glucose avec de la préculture pour obtenir une DO=0,05

Incubation 37°C 180rpm

- **Induction** à DO= 1.4 avec IPTG à 0,5mM

Incubation 20°C 150rpm overnight

- Cultures transvasées dans JA 8.1000
- Centrifugation 3500rpm 30min 4°C
- Resuspendre les culots dans 40ml PBS 1X , 0.5mM PMSF à 4°C
- Transvaser dans falcons 50ml
- Centrifugation 4500rpm 20min 4°C
- Culots conservés à -80°C

Purification du CTD de scMutL α

Matériel

⇒ Culots (2L) congelés à -80°C

Tampons

<u>Tp Lyse</u>	PBS 1 μ M phosphoramidon 1mM PMSF
----------------	---

<u>Tp Amylose</u>	PBS
0.22 μ M	1 μ M phosphoramidon 1mM PMSF

<u>Tp élution Amylose</u>	
0.22 μ M	PBS 1 μ M phosphoramidon 1mM PMSF 10mM Maltose

<u>Tp HisTrap</u>	20mM Na ₂ HPO ₄ pH7.4
0.22 μ M	500mM NaCl
pH ajusté	1 μ M phosphoramidon 1mM PMSF 20mM imidazole

<u>Tp HisTrap</u>	20mM Na ₂ HPO ₄ pH7.4
0.22 μ M	500mM NaCl
pH ajusté	1 μ M phosphoramidon 1mM PMSF 0.5M imidazole

<u>Tp A RessourceQ</u>	20mM Na ₂ HPO ₄ pH7.4
0.22 μ M	10mM β -mercaptoéthanol 1mM PMSF

<u>Tp B RessourceQ</u>	20mM Na ₂ HPO ₄ pH7.4
0.22 μ M	1M NaCl 1mM PMSF 10mM β -mercaptoéthanol

Colonnes/Résines

Amylose 35ml
HisTrap HP 5ml
Ressource Q 1ml B

Procédé

Lyse cellulaire

- ⇒ Le culot de 1L de culture est resuspendu dans 40ml de Tp de lyse.
- ⇒ Les cellules sont lysées par sonication (timer : 1min, température : 10°C, pulse : 1sec on et 1sec off, amplitude de 55% répété 5 fois)
- ⇒ Ajout Benzonase + MgCl₂ (10mM) incubation 15min à 4°C
- ⇒ Centrifugation à 12500rpm JA14 pendant 45min

Résine Amylose :

Equilibration :

- ⇒ 5CV de Tp Amylose

Chargement de la colonne (80ml)

- ⇒ A un débit de 2ml/min
- ⇒ Collection du flow through
- ⇒ Remettre dans Tp Amylose

Elution des protéines avec étiquette MBP (MBP-Pms1+ protéine GST-Mlh1 lié à MBP-Pms1)

5CV de Tp Amylose+maltose, élution collectée.

- Quantification des protéines de l'élution d'amylose par bradford

Coupure avec la TEV

- ⇒ Coupure de l'élution (dilué au 1/2) addition de 0.5% de TEV (0.6mg/ml) et incubation à 4°C overnight avec agitation très douce
- ⇒ Essayer de ne pas couper des fusions plus concentrées que 1-2mg/ml

Colonne d'affinité : HisTrap

Lavage et régénération de la colonne

- ⇒ 5CV de H₂O à 4ml/min
- ⇒ 5CV de 0.5M EDTA à 2ml/min
- ⇒ 5CV de H₂O à 4ml/min
- ⇒ 3ml de 0.1M NiSO₄ à 2ml/min
- ⇒ 5CV de H₂O à 4ml/min
- ⇒ 2.5ml de 0.1M NiSO₄ à 2ml/min

⇒ 5CV de H₂O à 4ml/min

Equilibration de la colonne

⇒ 5CV de Tp HisTrap

Chargement de la colonne (élution d'amylose coupée à la TEV) et élution de la protéine

⇒ A un débit de 2.5ml/min

⇒ Collection du flow through

⇒ Remettre dans Tp HisTrap

Elution des protéines avec étiquette his

⇒ 5CV de Tp HisTrap + imidazole, élution collectée

Quantification de notre complexe contenu dans le FT par nanodrop

Dilution au 1/10° de notre échantillon afin d'avoir une concentration en sel d'environ 50mM
Ou

Dialyse 3×2h contre le tampon 20mM Na₂HPO₄ pH7.4, 50mM NaCl, 10mM β-mercaptoéthanol, 1mM PMSF

Ressource Q 1ml:

- Equilibration avec le TpA: 20mM Na₂HPO₄ pH 7.4 , 10mM β-mercaptoéthanol, 1mM PMSF

- Chargement de notre complexe

- Gradient 5 à 100%B sur 20CV collecté par fraction de 2ml

- TpB: 20mM Na₂HPO₄ pH 7.4 , 1M NaCl, 10mM β-mercaptoéthanol, 1mM PMSF

- Débit: 1ml/min

ANNEXE 11: CONDITIONS DE CRISTALLOGENESE DU CRIBLE DU DOMAINE C-TERMINAL DE MLH1*PMS1

Score an nanti e	Req and numb er	Screen name	Salt	Buffer	pH	Precipitant 1	Precipitant 2	Note
03	A33	Crystal Screen- Nanite	0.1 M potassium chloride	0.25 M MES	5	10 %w/v PEG 400	0.01 M magnesium chloride hexahydrate	B
03	B03	Crystal Screen- MeritPlus	0.1 M lithium sulfate monohydrate	0.1 M Trisodium citrate dihydrate	5.8	12 %w/v PEG 6000		B
03	B03	Crystal Screen- Nanite	0.005 M magnesium ur sulfate	0.25 M MES	5	5 %w/v PEG 400		B
03	C12	Crystal Screen- Nanite	0.05 M ammonium acetate	0.25 M Tris hydrochloride	7.5	10 %w/v VPC		B
03	D07	Crystal Screen- Nanite	0.2 M potassium chloride	0.25 M MES	5.8	10 %w/v PEG 400	0.01 M magnesium chloride hexahydrate	B
03	D08	Crystal Screen- Nanite	0.1 M ammonium acetate	0.25 M sodium caprylate	6.5	10 %w/v 2-propanol	0.01 M magnesium sulfate	B
03	F38	Crystal Screen- Nanite	0.1 M potassium chloride	0.25 M sodium caprylate	5	15 %w/v 2-propanol	0.016 M magnesium acetate tetrahydrate	B
03	G07	Crystal Screen- Nanite	0.02 M magnesium chloride hexahydrate	0.25 M MES	5	15 %w/v 2-propanol	0.025 M magnesium chloride hexahydrate	B
03	G08	Crystal Screen- Nanite	0.04 M magnesium chloride hexahydrate	0.25 M sodium caprylate	5	5 %w/v MTD		B
06	A34	Peg RK		0.1 M HEPES	7	5 %w/v PEG 6000		B
06	B04	Peg RK		0.1 M HEPES	7	10 %w/v PEG 6000		B
06	B05	Peg RK		0.1 M Tris	9	10 %w/v PEG 6000		B
06	E04	VPC		0.1 M HEPES	7	10 %w/v 2-Methyl 2,4-pentenediol		B
06	E70	Screen- Mins 5000			7	5% Mins	0.1 M HEPES	B
06	B03	Screen- Index		0.1 M Bis-tris	5.5	25 %w/v PEG 3350		B
02	B06	Crystal Screen- I	0.2 M sodium iodide	20 %w/v PEG 3350	6.5	8 %w/v PEG 6000		B
03	E07	Crystal Screen- Nanite	0.2 M potassium chloride	0.25 M MES	5.8	5 %w/v PEG 6000	0.01 M magnesium chloride hexahydrate	B
03	F39	Crystal Screen- Nanite	0.03 M magnesium acetate tetrahydrate	0.25 M sodium caprylate	6.5	15 %w/v PEG 400		B
03	G09	Crystal Screen- Nanite	0.2 M potassium chloride	0.25 M sodium caprylate	6.5	10 %w/v PEG 4000	0.01 M magnesium chloride hexahydrate	B
03	G11	Crystal Screen- Nanite	0.1 M ammonium acetate	0.25 M HEPES Na	7	5 %w/v PEG 6000	0.02 M magnesium chloride hexahydrate	B
01	G11	Crystal Screen- II		0.1 M HEPES	7.5	70 %w/v VPC		4
01	E03	Crystal Screen- I	0.2 M magnesium acetate tetrahydrate	0.1 M sodium acetate	6.5	30 %w/v 2-methyl 2,4-pentenediol		4
01	E71	Crystal Screen- II		0.1 M HEPES	7.5	10 %w/v PEG 2000		4
01	H05	Crystal Screen- I		0.1 M Trisodium citrate dihydrate	5.8	20 %w/v 2-propanol	20 %w/v PEG 4000	4
02	A32	Crystal Screen- light	0.2 M ammonium acetate	0.1 M Trisodium citrate dihydrate	5.8	15 %w/v PEG 4000		4
02	A36	Crystal Screen- light		0.1 M Sodium I-HEPES	7.5	5 %w/v 2-propanol	10 %w/v PEG 4000	4
02	B03	Crystal Screen- light	0.05 M sodium dihydrogen phosphate	20 %w/v PEG 3350	6.8	10 %w/v PEG 2000		4
02	E08	Crystal Screen- PEG-300	0.2 M sodium thiocyanate					4
03	F34	Crystal Screen- MeritPlus	0.1 M sodium chloride	0.1 M HEPES Na	7.5	10 %w/v PEG 4000		4
06	A33	Peg RK		0.1 M MES	5	5 %w/v PEG 6000		4
06	E03	VPC		0.1 M MES	5	10 %w/v 2-Methyl 2,4-pentenediol		4
06	E09	Screen- Mins 5000			5	5% Mins	0.1 M MES	4
06	F35	VPC		0.1 M Tris	5	20 %w/v 2-Methyl 2,4-pentenediol		4
06	F36	VPC		0.1 M Bis-tris	5	20 %w/v 2-Methyl 2,4-pentenediol		4
06	F39	Screen- Mins 5000			5	10% Mins	0.1 M MES	4
06	F11	Screen- Mins 5000			5	10% Mins	0.1 M Tris	4
06	F12	Screen- Mins 5000			5	10% Mins	0.1 M Bis-tris	4
06	C07	Screen- Index	0.2 M ammonium acetate	0.1 M Bis-tris	6.5	45 %w/v 2-methyl 2,4-pentenediol		4
06	E07	Screen- Index	0.2 M ammonium acetate	0.1 M HEPES	7.5	45 %w/v 2-methyl 2,4-pentenediol		4
06	E07	Screen- Index	0.2 M ammonium acetate	0.1 M Tris	6.5	45 %w/v 2-methyl 2,4-pentenediol		4
06	G08	Screen- Index	5.00 %w/v lactamide pH 7.0	0.1 M HEPES	7	10 %w/v PEG VME 5.00		4
02	F33	Crystal Screen- light	0.2 M sodium acetate trihydrate	0.1 M Tris hydrochloride	8.5	15 %w/v PEG 4000		3
02	F12	Crystal Screen- PEG-300	0.2 M Trisodium citrate dihydrate	20 %w/v PEG 3350	6.2			3
03	G12	Crystal Screen- Nanite	0.1 M potassium chloride	0.25 M Tris hydrochloride	8.5	30 %w/v PEG 400	0.01 M magnesium chloride hexahydrate	3
06	B03	Peg RK		0.1 M MES	5	10 %w/v PEG 6000		3
06	F10	Screen- Mins 5000			7	10% Mins	0.1 M HEPES	3
06	G05	VPC		0.1 M Tris	8	40 %w/v 2-Methyl 2,4-pentenediol		3
06	G06	VPC		0.1 M Bis-tris	9	40 %w/v 2-Methyl 2,4-pentenediol		3
06	A12	Screen- Index	0.10 M sucrose pH 7.0		15	%w/v PEG 3350		3
06	B03	Screen- Index		0.1 M Bis-tris	6.5	45 %w/v 2-methyl 2,4-pentenediol		3
06	E03	Screen- Index	0.20 M sucrose	0.1 M HEPES	7.5	10 %w/v PEG 3350		3

En rouge = Conditions ayant données des cristaux

En violet = Conditions ayant données des microcristaux, des sphérulites ou du précipité cristallin

ANNEXE 12: SCRIPTS DE REMPLACEMENT MOLECULAIRE

Remplacement moléculaire de deux monomères avec MolRep

```
#!/bin/sh

#PBS -l nodes=1:ppn=4
#PBS -l walltime=600:00:00
#PBS -N mol2mod*
#PBS -j oe
#PBS -o molrep_loop2oe

source /share/apps/buster/setup.sh
REFINE="/share/apps/buster/autoBUSTER/bin/linux/refine"

cd /home/lbsr/charbo/Emeric2/molrep_loop2/

np=30
aniso="Y"
score="N"

models=$( ls /home/lbsr/charbo/Emeric2/*.pdb )
echo $models
for model in $models ; do
  for resmax in 3 4 5 6 7 ; do
    model_name=$(basename $model .pdb)
    echo "Modele:" $model_name
    echo "Resolution=" $resmaxho

cat << eof > molrep.inp
RESMAX $resmax
ANISO $aniso
MODE F
NP $np
SCORE $score
NMON 1
SURF A
eof

#Remplacement moléculaire d'un monomère
molrep -k molrep.inp -m $model -f
/home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz >
molrep1_${model_name}_${resmax}.log
mv molrep.pdb molrep1_${model_name}_${resmax}.pdb

#Corps rigide du 1er monomère issu du remplacement moléculaire
$REFINE -p molrep1_${model_name}_${resmax}.pdb \
  -m /home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz \
  -d buster1_${model_name}_${resmax} \
  -nthreads 4 \
  -nbig 10 \
  -RB \
  -autoncs_noprune \
  StopOnGellySanityCheckError=no \
  > buster1_${model_name}_${resmax}.log
```



```

rm -r
/home/lbsr/charbo/Emeric2/molrep_loop2/buster1_${model_name}_${resmax}/01-
BUSTER

#Remplacement moléculaire d'un second monomère
molrep -k molrep.inp -m $model -mx
buster1_${model_name}_${resmax}/refine.pdb -f
/home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz >
molrep2_${model_name}_${resmax}.log
mv molrep.pdb molrep2_${model_name}_${resmax}.pdb

#Affinement de 2 monomères issus du remplacement moléculaire
$REFINE -p molrep2_${model_name}_${resmax}.pdb \
        -m /home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz \
        -d buster2_${model_name}_${resmax} \
        -nthreads 4 \
        -nbig 10 \
        -RB \
        -autoncs_noprune \
        StopOnGellySanityCheckError=no \
        > buster2_${model_name}_${resmax}.log

rm -r
/home/lbsr/charbo/Emeric2/molrep_loop2/buster2_${model_name}_${resmax}/01-
BUSTER

#Affinement de 2 monomères issus du remplacement moléculaire dans le jeu
SeMeth
refmac5 hklin /home/lbsr/charbo/Emeric2/MTZ/xds_m04-08_pk_se.mtz \
        xyzin buster2_${model_name}_${resmax}/refine.pdb \
        hklout rigid_${model_name}_${resmax}.mtz \
        xyzout rigid_${model_name}_${resmax}.pdb << eof >
rigid_${model_name}_${resmax}.log
        refine type rigid
        rigidbody ncycle 10
        rigidbody group 1 from 528 A to 757 A
        rigidbody group 2 from 528 B to 757 B
        monitor medium
eof

#Recherche de pic anomaux dans le jeu SeMeth (script issu de P. Legrand-
Synchrotron Soleil)
rm -f *.geo *.eff
sh run Phaser.sh toto rigid_${model_name}_${resmax}.pdb

done
done

```

Remplacement moléculaire de deux monomères avec Phaser

```

#!/bin/sh

#PBS -l nodes=1:ppn=4
#PBS -l walltime=600:00:00
#PBS -N phas6mod*
#PBS -j oe
#PBS -o pharep_loop6.oe

source /share/apps/buster/setup.sh

```

```

REFINE="/share/apps/buster/autoBUSTER/bin/linux/refine"

cd /home/lbsr/charbo/Emeric2/phaser_loop6/

np=30
aniso="Y"
score="N"

models=$( ls /home/lbsr/charbo/Emeric2/*.pdb )
echo $models
for model in $models ; do
  for resmax in 3 4 5 6 7 ; do
    model_name=$(basename $model .pdb)
    echo "Modele:" $model_name
    echo "Resolution=" $resmax

#Remplacement moléculaire de deux monomères
phaser << eof > mrphaser1_${model_name}_${resmax}.log
TITLE mrphaser_${model_name}_${resmax}
MODE MR_AUTO
HKLIn /home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz
LABIn F=F_Native SIGF=SIGF_Native
RESOLution 40 $resmax
ENSEmble ${model_name} PDB $model IDENTity 20
COMPosition PROTein MW 61000 NUM 2 #2monomères
SEARCh ENSEmble ${model_name} NUM 2
FINAL SELECT PERCENT 65
ROOT mrphaser1_${model_name}_${resmax}
eof

#Affinement de 2 monomères issus du remplacement moléculaire
$REFINE -p mrphaser1_${model_name}_${resmax}.1.pdb \
-m /home/lbsr/charbo/Emeric2/MTZ/INTEGRATE_Native_sort180.mtz \
-d buster2_${model_name}_${resmax} \
-nthreads 4 \
-nbig 10 \
-RB \
-autoncs_noprune \
StopOnGellySanityCheckError=no \
> buster2_${model_name}_${resmax}.log

rm -r
/home/lbsr/charbo/Emeric2/molrep_loop1/buster2_${model_name}_${resmax}/01-
BUSTER/

#Affinement de 2 monomères issus du remplacement moléculaire dans le jeu
SeMeth
refmac5 hklin /home/lbsr/charbo/Emeric2/MTZ/xds_m04-08_pk_se.mtz \
xyzin buster2_${model_name}_${resmax}/refine.pdb \
hklout rigid_${model_name}_${resmax}.mtz \
xyzout rigid_${model_name}_${resmax}.pdb << eof >
rigid_${model_name}_${resmax}.log
  refine type rigid
  rigidbody ncycle 10
  rigidbody group 1 from 433 A to 603 A
  rigidbody group 2 from 433 B to 603 B
  monitor medium
eof

#Recherche de pic anomaux dans le jeu SeMeth (script issu de P. Legrand-
Synchrotron Soleil)

```

```
rm -f *.geo *.eff
sh run Phaser.sh toto rigid_${model_name}_${resmax}.pdb

done
done
```

ANNEXE 13: COMPARAISON DES STRUCTURES SECONDAIRES PREDITES ET REELLES DES CTD DE MLH1 ET PMS1

Comparaison des structures secondaires prédites par Jpred et celles de la structure cristallographique du CTD de Mlh1

Mlh1 <i>S. cerevisiae</i>	VNLTSIKKLREKVDDSIHRELTDIFANLNYGVVDEERRLAAIQHDLKFLIDYGSVCYELF
Prédiction (Jpred)	BB--HHHHHHHHHHHH--HHHHHHHH--BBBBB--BBBBBBB--BBB--HHHHHHHHH--
Structure	--HHHHHHHHHHHH--HHHHHHHH--BBBBBBB--BBBBBBB--BBBBBBHHHHHHHHH--
Mlh1 <i>S. cerevisiae</i>	YQIGLTDIFANFGKINLQSTNVSDDIVLYNLLSEFDELNDASKEKIISKIWDMSMLNEYYSIELV
Prédiction (Jpred)	HHHHHH--B--HHHHHH--HHHHHHHHHHHHHHHHHH--BBBB
Structure	HHHHHH--BBB--HHH--HHHHHH--HHHHHHHHHHHHHHHHHH--BBB
Mlh1 <i>S. cerevisiae</i>	NDGLDNDLKSVKLKSLPLLLLKGYIPSLVKLPFFIYRLGKEVDWEDEQECLDGLIREIALLYIPDMV
Prédiction (Jpred)	-----BBBBBB-----HHHHHH--HHHHHHHHHHHH--
Structure	B-----BBBBBB-----HHHHHHHHHHHH--HHHHHHHHHHHH--
Mlh1 <i>S. cerevisiae</i>	PKVDTSASLSSEDEKAQFINRKEHISLLEHVLFPCIKRRFLAPRHILKDVVEIANLPDLYKVFERC
Prédiction (Jpred)	-----HHHHHHHHHHHHHHHHHHHH--HHHHH--BBBBBB-----
Structure	-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHBBBBBBHHHHHHHH--

Comparaison des structures secondaires prédites par Jpred et celles de la structure cristallographique du CTD de Pms1

Pms1 <i>S. cerevisiae</i>	ELEDFEQGEKYLTLTVSKNDFKKMEVVQGPNLGFIIIVTRKVDNKYDLFIVDQHASDEKYNFETLQAVT
Prédiction (Jpred)	-----BBBBBB--BBBBBBB--BBBBB--HHHHHHHHHHHHHH--
Structure	--HHHHHHHHHHHH--HHHHHH--BBBBBB--BBBBBBB--BBBBBBBHHHHHHHHHHHHHH--
Pms1 <i>S. cerevisiae</i>	VEKSQKLIIPQVVELSVIDELVVLNLFVFEKNGFKLKIDEEEFGSRVKLLSLPTSQQTFLDGLGFN
Prédiction (Jpred)	-----BBB--HHHHHHHHHHHHHHHHHH--BBBBB--BBBBBBB--HHHHHH
Structure	---BBB--BBBBBB-----HHHHHH--HHHHHH--BBBBBB--BBBBBBB--HHHHHH
Pms1 <i>S. cerevisiae</i>	ELIHLIKEDGGLARDNIRCSKIRSMFAMRACRSSIMIGKPLNKKTMTRVVHNLSELDKPWNCPHGRPT
Prédiction (Jpred)	HHHHHHHH--HHHHHHHHHHHHHHHHHH--HHHHHHHHHHHH--BB
Structure	HHHHHHHH--HHHHHHHHHHHHHHHHHH--HHHHHHHHHHHH--
Pms1 <i>S. cerevisiae</i>	MRHLMELRDWSSF SKDYEI
Prédiction (Jpred)	BBBB--HHHHHHHH--
Structure	BBBBBB-----

ANNEXE 14: ANALYSE DE L'INTERFACE DE DIMERISATION PAR LE SERVEUR PROTORP ; RESIDUS DE MLH1 A LA SURFACE DE DIMERISATION

Résidu	Surface accessible (Å ²)	Surface accessible (%)	Résidu	Surface accessible (Å ²)	Surface accessible (%)
N537	2.74	0.15	R672	47.93	2.65
V539	76.47	4.23	K675	1.05	0.06
G540	15.15	0.84	E676	24.21	1.34
V541	19.49	1.08	R691	34.50	1.91
V542	80.57	4.46	L695	34.19	1.89
D543	23.57	1.30	I698	10.48	0.58
E544	10.03	0.55	D700	3.93	0.22
R546	118.47	6.55	M701	62.24	3.44
R547	28.21	1.56	K704	78.62	4.35
L548	42.97	2.38	D706	27.36	1.51
A550	7.73	0.43	S708	11.39	0.63
Q552	36.89	2.04	K724	36.81	2.04
D554	3.39	0.19	E725	25.00	1.38
K556	1.54	0.09	K751	4.78	0.26
L557	51.63	2.86	D752	12.90	0.71
L559	38.88	2.15	V754	21.11	1.17
D561	36.31	2.01	I756	61.07	3.38
G563	5.58	0.31	A757	12.28	0.68
S564	18.76	1.04	L759	43.01	2.38
Y567	21.25	1.18	L762	43.70	2.42
E568	1.38	0.08	Y763	40.85	2.26
F604	8.90	0.49	V765	77.32	4.28
V664	5.83	0.32	F766	133.38	7.38
K665	29.04	1.61	E767	47.37	2.62
F668	49.75	2.75	R768	9.89	0.55
Y671	32.70	1.81	C769	135.00	7.47

NB : les résidus marque la frontière entre les différents segments de l'interface de dimérisation

ANNEXE 15: ANALYSE DE L'INTERFACE DE DIMERISATION PAR LE SERVEUR PROTORP ; RESIDUS DE PMS1 A LA SURFACE DE DIMERISATION

Résidu	Surface accessible (Å ²)	Surface accessible (%)	Résidu	Surface accessible (Å ²)	Surface accessible (%)
D654	8.24	0.46	C848	3.88	0.21
F655	26.62	1.47	H850	21.39	1.18
K673	46.72	2.58	R852	77.85	4.31
E675	33.05	1.83	P853	39.54	2.19
V677	78.39	4.34	T854	1.03	0.06
G678	14.95	0.83	M855	3.32	0.18
Q679	18.83	1.04	R856	26.29	1.45
F680	109.73	6.07	L858	51.85	2.87
N681	73.82	4.08	M859	39.05	2.16
G683	1.88	0.10	L861	36.44	2.02
F684	69.47	3.84	R862	18.80	1.04
I686	23.43	1.30	T864	148.61	8.22
T688	9.48	0.52	S866	58.32	3.23
K690	70.12	3.88	F867	124.54	6.89
D692	9.76	0.54	S868	12.16	0.67
N693	4.51	0.25	K869	58.72	3.25
Y695	65.56	3.63	D870	104.58	5.79
L697	30.71	1.70	Y871	128.56	7.11
I699	29.21	1.62	E872	16.95	0.94
D701	24.21	1.34	I873	101.61	5.62
H703	13.75	0.76			
A704	3.20	0.18			
E707	1.73	0.10			
C817	12.08	0.67			
I823	4.91	0.27			
K829	24.24	1.34			

NB : les résidus marque la frontière entre les différents segments de l'interface de dimérisation

ANNEXE 16: ALIGNEMENTS DE 10 SEQUENCES DU DOMAINE C-TERMINAL DE MLH1 A L'AIDE DU PROGRAMME KALIGN

MLH1	<i>A. thaliana</i>	453	DLSSVQELIAGVDSCHPGMLETVRNCTYVG--MADD--VFALVQYNTHLYLANVVNLSKELMYQQTILRRFAHFN-----AIQI-SDPAPLSELILL--A	540
MLH1	<i>D. rerio</i>	476	KLTSIKGLRDQIELQTHKGLQELLQNHSEFVG--SVSP--QWTLVQHQTCLYLLNTTKLSQELFYQILIYDFGNFG-----VLRL-SNPAPLYDLAML--A	563
MLH1	<i>D. discoideum</i>	622	ELTSIKTLISEFQSNVHDGLKEFFNDCVFVG--CLDH--SYALVQFGKKLYLINLENITKELFYQLSLRRFSDFD-----SIKF-SQSLSIYSILLV--S	709
MLH1	<i>H. sapiens</i>	502	NLTSVLSLQEEINEQGHEVLREMLHNHSFVG--CVNP--QWALAQHQTCLYLLNTTKLSEELFYQILIYDFANFG-----VLRL-SEPAPLFDLAML--A	589
MLH1	<i>D. persimilis</i>	437	RLSSVLDMRRSVERQCNCVQLRSILKNLVYVG--CVDE--RRALFQHETRLYLCTNTRAFSEELFYQRLVYEFQNC-----EISV-MPPLSLHELLMI--A	524
MLH1	<i>S. pombe</i>	447	DLESIKSLQKQVINSMHVLTATNLTETHKYVGLVCPTR--RIAAVQHNIIGLYVVDYGKLSYHLYFIQICLTFEGNYG-----EFVL-ETPLSISDLFEI--V	536
MLH1	<i>S. cerevisiae</i>	510	NLTSIKKLREKVDSDSIHRELTDFANLNYVG--VVDEERRLAAIQHDLKFLIDYGSVCYELFYQIGLTFANFG-----KINL-QSTNVSDDIVLY--N	599
MLH1	<i>C. elegans</i>	501	HFESIEVLRKEIIANSSQSLREMFKTSTFVG--SINV--KQVLIQFGTSLYHLDFTSLVREFFYQISVFSFGNYG-----SYRLDEEPPAIIEILELLGE	591
MLH1	<i>L. major</i>	932	LLSSVSMIVDRLLAEASPTADNLVDQLSEFVG--TVDS--RAFLAQAGTTLWCDTALTRHVVVFQRIFLRWCQPALPAPPVLAF-ATPVRLADLLLL--A	1024
MLH1	<i>P. falciparum</i>	710	EISSIKKLKMCIEEKEKKELETECLKNSIYVG--PVDN--MHSLIQYKEKLLLIKMPLIKEVITYQSILNRLGKIP-----PFEF-DPPIPLYDLLLV--A	797
MLH1	<i>A. thaliana</i>	541	LK-----EEDLDPGNDTK--DDLKERIA-----	561
MLH1	<i>D. rerio</i>	564	LD-----SEE--SGWTEE--DGPKEGLA-----	582
MLH1	<i>D. discoideum</i>	710	LD-----SPL--SGWME--DGPDKIA-----	728
MLH1	<i>H. sapiens</i>	590	LD-----SPE--SGWTEE--DGPKEGLA-----	608
MLH1	<i>D. persimilis</i>	525	LD-----SEA--AGWTPE--DGDKTELA-----	543
MLH1	<i>S. pombe</i>	537	-----N--GDEKSES-----	545
MLH1	<i>S. cerevisiae</i>	600	LL-----SEF-----D--ELNDDASK-----	613
MLH1	<i>C. elegans</i>	592	LS-----TRE--PNYAALF--EVFANVEN-----	610
MLH1	<i>L. major</i>	1025	LAYDGPLHQPPTSATLLAVVEEAKKRQQQQAQVAGGGAAQLATSSLLTDVVRQTGGATA--RAWREV-LF-NADMESKTLSPNDLHTESGAGGVFAGSES	1120
MLH1	<i>P. falciparum</i>	798	VN-----NSY--SGFYENPNYANKNIEKVC-----	820
MLH1	<i>A. thaliana</i>	562	-----EMNTE-LLKEKAEMLEEFVSVHID--SSA-----NLSRLPV-ILDQYTPDMDRVPEFLLCIGND-VEW-----	619
MLH1	<i>D. rerio</i>	583	-----QYIVD-FLKQKAEMLEEFVSLIED--AEG-----NLTLGLPM-LLDNYTPAMEGLPMFILRLATE-VNW-----	640
MLH1	<i>D. discoideum</i>	729	-----DYLTK-LLISKKELNLEYFSIEIN--EDG-----ELVGIPQ-VLDHVVPCDNLPIFLKLATE-VEW-----	786
MLH1	<i>H. sapiens</i>	609	-----EYIVE-FLKKAEMLADYFSLEID--EEG-----NLIGLPL-LIDNVVPPLEGLPIFILRLATE-VNW-----	666
MLH1	<i>D. persimilis</i>	544	-----ASAVQ-ILQOKAPIMKEYFGLRIS--EEG-----LVESLPS-LVGRHKPSATHLPVYLLRLATE-VDW-----	601
MLH1	<i>S. pombe</i>	546	-----EKFTR-LLVSRDMLKDYFSISVT--SGG-----LLTAVPM-LSPKYHPPFEQLPLLISSLTPKFFDW-----	604
MLH1	<i>S. cerevisiae</i>	614	-----EKIIS-KIWDMSMLNEYYSIELV--NDGLDNDLKSVKLSLPL-LLKGYIPSLVKLPFFIYRLGKE-VNW-----	679
MLH1	<i>C. elegans</i>	611	-----RFAAEKLLAEHADLLHDYFAIKLDQLENG--RL--HTEIPS-LVHYFVPQLEKLPFLIATLVLN-VDY-----	673
MLH1	<i>L. major</i>	1121	VLPQPEDATMRYVRRLVR-RLCRWRGLLKEYFYIDIT--ADG-----LLVGLPYGLNRHWPMPRAVPMVWLLAE-VPYPGTAAPSSSTMSGAAD	1208
MLH1	<i>P. falciparum</i>	821	-----NELEQ-IFYTYNEMYSDFYSIII--EDG-----CIVTFPA-CCGEYFGQEFPLPFLFLRLATQ-IDY-----	877

MLH1	<i>A. thaliana</i>	620	-----EDEKSCFQGVSAAGN-FYAMHPPLLNPSSGDGIQFYSKRG	659
MLH1	<i>D. rerio</i>	641	-----DKEKECFREFSVESCH-FYSIRKSYT-----	665
MLH1	<i>D. discoideum</i>	787	-----EFEKECFAGIVKEISS-FFKIEPSFL-----	811
MLH1	<i>H. sapiens</i>	667	-----DEEKECFESLSKECAM-FYSIRKQYI-----	691
MLH1	<i>D. persimilis</i>	602	-----EQEAKCFESFCRETAR-FYAQQDWH-----	626
MLH1	<i>S. pombe</i>	605	-----LDEKSCNLNGIMKAIK-FYVPLPLSY-----	629
MLH1	<i>S. cerevisiae</i>	680	-----EDEQECLDGLREIAL-LYIPDMVLK-----	704
MLH1	<i>C. elegans</i>	674	-----DDEQNTFRTICRAIGD-LFTLDTNFI-----	698
MLH1	<i>L. major</i>	1209	QTGPTSAAWQSNENRGAEVVCVATSTSPATSLAPTQTEATPTTTPQPTAMEAEADGVAQEVACFTAVARHIAETLYGLPPPPP-----STPP	1297
MLH1	<i>P. falciparum</i>	878	-----SKEINCINGICYLLAN-FYSKITLLN-----DKEWYQDDL	913

MLH1	<i>A. thaliana</i>	660	ESSQEKSDLEGNVDMEDN--L--DQDLLSDAENAW-----AQR--EWSI-----QHVLFPSMR--LFLKPPASMASN	718
MLH1	<i>D. rerio</i>	666	-----LEA-----DADEPQDAEMSW-----QWKV-----EHVLFKALR--SLFSPAKHLS	705
MLH1	<i>D. discoideum</i>	812	-----KLR-----DTQVNNQQQTNSTTTNNINFIKKGKEWII-----QHLIFPAFR--K-LSPPKKFAND	865
MLH1	<i>H. sapiens</i>	692	-----SEESTLSGQQSEVPGSIPNSW-----KWTV-----EHIVYKALR--SHILPPKHFT	737
MLH1	<i>D. persimilis</i>	627	-----DES-----SCHLP-----WQT-----EHVLFPAK--KYLVPARLRQ	658
MLH1	<i>S. pombe</i>	630	-----EES-----DVKSIRSLESC-----L-----EDYLFPEFR--RRVICPKKVFE	665
MLH1	<i>S. cerevisiae</i>	705	-----V--DTS-----DASLSEDEKAQF-----INRKEHI--SSLL-----EHVLFPCIK--RRFLAPRHILKD	752
MLH1	<i>C. elegans</i>	699	-----TL--DKKISAFSATPW-----KTLI-----KEVLMPLVK--RKFIPEEHFKQA	737
MLH1	<i>L. major</i>	1298	TSAEGP-----DTA-----TADVAAGATDTA--SAE-----LWRE-----ELV-RHGLFACLKNPQLCRLPDQCLRD	1351
MLH1	<i>P. falciparum</i>	914	MIQEKEKETQMLL--NKS-----NKHINNSNHHNQ-----HYDETNL-DYILGDESVVDINKHLSVSRNINLVFEKYFFPMIQLNNIMKIPTTFSNN	997

MLH1	<i>A. thaliana</i>	719	GTFVKVASLEKLYKIFERC--	737
MLH1	<i>D. rerio</i>	706	GSLVLIASLPDLYKVFERC--	724
MLH1	<i>D. discoideum</i>	866	GSVIQTITLDNLYKVFERC--	884
MLH1	<i>H. sapiens</i>	738	GNILQLANLPDLYKVFERC--	756
MLH1	<i>D. persimilis</i>	659	--IYELTNLPTLYKVFERC--	675
MLH1	<i>S. pombe</i>	666	KCIYQITSLPRLYNVFERC--	684
MLH1	<i>S. cerevisiae</i>	753	--VVEIANLPGLYKVFERC--	769
MLH1	<i>C. elegans</i>	738	GVIQLADSHDLYKVFERCGT	758
MLH1	<i>L. major</i>	1352	GTIQSLVSVESLYKVFERC--	1370
MLH1	<i>P. falciparum</i>	998	GYIIELTSLNQLYKIFERC--	1016

ANNEXE 17: ALIGNEMENTS DE 10 SEQUENCES DU DOMAINE C-TERMINAL DE PMS1 ET 2 SEQUENCES DU DOMAINE C-TERMINAL DE MLH3 A L'AIDE DU PROGRAMME KALIGN

Pms1	<i>A. thaliana</i>	659	TPQP K CFAAATLELSQP D DEERKARALAAATSE L ERLFRKEDFRMQVLGQFN L GFIIAKL-----ERD-----	723
Pms1	<i>D. rerio</i>	627	SNEQ E PKYRRFRAKIN P GENQTA-----EDEL K KEISKDMFK E MEIIGQFN L GFII T KI-----KSD-----	683
Pms1	<i>D. discoideum</i>	776	QKAATQVTSQLQQQPSQT N QKTA-----EEEL T KFFK E YFKQ M IVIGQFN L GFIIAKL-----GND-----	832
PMS2	<i>H. sapiens</i>	638	QSEGEQNYRK F RAKIC P GENQAA-----EDEL R K E ISK T MF A EMEIGQFN L GFII T KL-----NED-----	694
Pms1	<i>S. pombe</i>	585	-----GVV H DISDENQED H -----LNL T V H KADFLRM R VVGQFN R GFIVV V H-----GNN-----	629
Pms1	<i>A. niger</i>	585	RRWEE H HG P SQSEDETAAPQ E TA-----EERLS L TVSKDDFAN M RIVGQFN L GFIL A TR-----SSTDDSG P GSSAKDE-----	653
Pms1	<i>S. cerevisiae</i>	636	KRKSE A QENI I KNKDELEDFEQ G -----EKY L TLTVSKND F K K MEVVGQFN L GFII V TR-----KVDN-----KYD-----	696
Pms1	<i>C. elegans</i>	581	VSAEKDVLNEITTKINKEEN D DA-----ERQ L SRSLTKDD F SK M KIIGQFN H GFII C RL-----RGH-----	637
Pms1	<i>L. major</i>	621	QTCDRPSSRTLGAQTA-----DDL N HYFTKNS F K E MRVIGQFN H GFII A VL-----P-----NGD-----	670
Pms1	<i>P. falciparum</i>	962	SNNNTN N VNNNEDIN F SNID E T-----Q K DL--YFQSS L FN K L K ICGQFN K GFV S ISKID L LY F E K KK K Y-----GNEG H ES E CY K TH D NN S NNI	1045
MLh3	<i>S. cerevisiae</i>	459	KLSSRPYDSGF T EDYDSIG K IT I T-----D F SI S RSVLAKYEVINQ V DK K FIL I RC--L-----DQS-----I H -----	514
MLH3	<i>H. sapiens</i>	1146	DNPVFARYPEVAVDVSSGQ A ESL-AVK I H--N I LYPYR F T K G M I H SMQVLQ V DN K FI A CLM-----ST K TE-----E-NGE-----	1213

Pms1	<i>A. thaliana</i>	725	-----LFIVDQHA A DEK F N F EH L -----AR S T--V-----LNQQ P LL-QPL N LELS P EE V TVLM H DI I RE N G F LL E -----	784
Pms1	<i>D. rerio</i>	684	-----LFIIIDQH A TDEK Y N F EM L -----QQ N T--V-----L K G Q RLI-VP Q SL H PA S ETV L ME N LE I FR K NG F DL I ---	744
Pms1	<i>D. discoideum</i>	833	-----LFIIIDQH A DEK Y N F E I L-----SK S VE S S-----INS Q PL K PD T LS D LT S EE E LI I IE N VD L FK K NG F K I I---	696
PMS2	<i>H. sapiens</i>	695	-----IFIVDQH A TDEK Y N F EM L -----QQ H T--V-----L Q G Q RLI-AP Q TL N L T AV N E A V L IE N LE I FR K NG F DF V I---	755
Pms1	<i>S. pombe</i>	630	-----LFIIIDQH A SDEK F N Y EH L -----K S N L --V-----INS Q DLV-L P K R LD L A A TE E TV L ID H ID L IR R K G F G VA I ---	690
Pms1	<i>A. niger</i>	654	-----LFIIIDQH A SDEK F N F ER L -----QA E T--V-----V Q N Q RLV-Q P K R LD L T A VE E E I VI E N Q A A LE K NG F IV E V---	714
Pms1	<i>S. cerevisiae</i>	697	-----LFIVDQH A SDEK Y N F ET L -----Q A V T --V-----F K S Q KLI-IP Q PV E LS V ID E LV V LD N L P V F E K NG F K L K I ---	757
Pms1	<i>C. elegans</i>	638	-----LFIVDQH A SDEK Y N F ER L -----Q S SA-- K -----L T K Q PL F -M P T A L G F G AV Q EL I IRE N L P I F H A NG F DF E F---	698
Pms1	<i>L. major</i>	671	-----VFVVDQH A SDEK Y N Y ER L -----V R AY-- E -----AT P Q P LV-M P V S V A MS A HE V DL A VE H K L AL Q H H G F K V S R ---	730
Pms1	<i>P. falciparum</i>	1046	NCEDYDN F SN D N K HS N YALFIIDQH A DEK S N F E K Y-----N K I F --T-----M K S Q KLI-S K I H V Q V S PA Q V H I Q K Y MS I FL Q NG F EV Q ILE E	1128
MLh3	<i>S. cerevisiae</i>	515	NCPL-----LVLVDQH A CDER I RLEEL F Y S LL T EV T G--T-----F V ARDL--K D CCIE V DR T E A DL F K H Y Q SE F K K W G I G Y E T---	585
MLH3	<i>H. sapiens</i>	1214	-----AGG N L-----LVLVDQH A A H ER I RLE Q L-I D SY- E K Q Q--AQGS G R K KL S ST L I-PPLE I TV T EE Q RRLLW C Y H K N LE D L G LE F V F ---	1291

Pms1	<i>A. thaliana</i>	785	-----NP--S A PPG K H F RL R AI P Y-----SK N IT F -G V ED L K D L I ST L G D -----N H GE C S-----V A SS Y K T SK T D	838
Pms1	<i>D. rerio</i>	744	-----DE--DAQ V MD R V K L V SL P T-----SK N WT F -G P ND I EEL I F M LS D -----S P G I MC-----R-----	788
Pms1	<i>D. discoideum</i>	696	-----DH--D A PT R F K IK L SA F PI-----I H G Q S F -G I K D Y E W I F M I K E-----SS I P G -----S V N	938
PMS2	<i>H. sapiens</i>	756	-----DE--N A PV T ER A K L IS L P T -----SK N WT F -G P Q D VE L IF M LS D -----S P G V MC-----R-----	799
Pms1	<i>S. pombe</i>	691	-----DL--N Q RV G N R CT L LS V PT-----SK N V I F-D T SD L LE I IS V LS E -----H-----P Q ID	732
Pms1	<i>A. niger</i>	715	-----DESGDE P IG R R C K L VS L PL-----S K E V V F -G V R D LEEL I V L LS E T P T N A A RS A T G M Y I P R-----	769
Pms1	<i>S. cerevisiae</i>	758	-----DE--E E E F GS R V K LL S L P T-----S K Q T LF-D L G D F N EL I H L I K E-----D G GL R RD N I-----R-----	804
Pms1	<i>C. elegans</i>	699	-----SE--N D GC I -K T FL T AR E -----LL N Q Q L-T N SD L EE I L A V V S Q -----Y P N Q MY-----R-----	741
Pms1	<i>L. major</i>	731	-----GS--D D --T K LL V YS L P V -----L P YDVV-S A S D V M ELV Q QL V Q-----Y G T I T K -----P-----	772
Pms1	<i>P. falciparum</i>	1129	PI H K R R K T N NN N IN E PI D DEE--E M L M EL N V Y LL S LP V -----F N G K IL-E V VD F MS L L H HL E --H P V A S Y -----N--E S EV S V K	1198
MLh3	<i>S. cerevisiae</i>	586	-----I E --G T M E TS L LE I K T LP E -----M L T S K Y NG D K D Y L K M V L L Q H A -----H D L K D F -----K-K-L P MD L SH F E	640
MLH3	<i>H. sapiens</i>	1292	-----P D TS D SL V LV G K V PL C F V ERE A N E LR R GR S T V TK S I-V E EF I RE Q LE L L Q T--T G G I Q G -T L -----	1349

Pms1	<i>A. thaliana</i>	839	SICPSRVRA-----MLASRACRSSVMIGDPLRKNEQKIVEHLADLESPPWNCPHGRPTMRHLV-DLTTLTLTPDDDNVNDDDDDDAT	919
Pms1	<i>D. rerio</i>	789	---PSRVQR-----MFASRACRKSVMTGTALNTSEMKKLVLMGEIEQPWNCPHGRPTMRHLA-NL-DMISQD-----	851
Pms1	<i>D. discoideum</i>	939	K--IPRLNS-----LLASKACRKSIMVGTTLTHKEMKDVLNNLSTLDNPWCPCPHGRPTMRHLV-DL-SIKDKL-KQQQQQQQKQQQ	1018
PMS2	<i>H. sapiens</i>	800	---PSRVKQ-----MFASRACRKSVMTGTALNTSEMKKLITHMGEMDHPWNCPHGRPTMRHLA-NL-GVISQN-----	862
Pms1	<i>S. pombe</i>	733	PF-SSRLER-----MLASKACRSSVMIGRALTISEMNTIVRHAELESKPWNCPHGRPTMRHLL-RLKDI-----	794
Pms1	<i>A. niger</i>	770	---PSKVRK-----MFAMRACRSSIMIGKTLTVKQMQVVQNMGTIDKPWNCPHGRPTMRHLM-SLGQWNEWD---ELDDEEDQPV	843
Pms1	<i>S. cerevisiae</i>	805	---CSKIRS-----MFAMRACRSSIMIGKPLNKKTMTRVVHNLSELDKPWNCPHGRPTMRHLM-ELRDWSSFS-KDYEI-----	873
Pms1	<i>C. elegans</i>	742	---PVRIRK-----IFASKACRKSVMTGKPLNQREMTQIIRHLAKLDQPWNCPHGRPTIRHLA-SLPDRAEFE-----	805
Pms1	<i>L. major</i>	773	---LRAVWH-----SMATKACRSSIMIGTPLTVKRMKLILERLSQLDQPWNCPHGRPTLRLLC-NIVDLSRGG---ALLR-----	840
Pms1	<i>P. falciparum</i>	1199	T--TIDLNNKTDTWFNYNFPRPQKVWRILASKACRNAIMVGKALNIYEMIKIKKLSFLKNPWNCPHGRPTIKYLINNV-DIKNCF-KNYYLKLYDEITN	1294
Mlh3	<i>S. cerevisiae</i>	641	N--YTSVDK--LYWWKYSSCVPTVFHEILNSKACRSVMFGDELTRQECIILISKLSRCHNPFECAGHRPSMVPIA-ELK-----	715
MLH3	<i>H. sapiens</i>	1350	---PLTVQK-----VLASQACHGAIKFNDGLSLQESCRLLIEALSSCQLPFQCAHGRPSMLPLA-DIDHLEQEK---QIKPNLTCLR	1423

Pms1	<i>A. thaliana</i>	920	ISLA-----	923
Pms1	<i>D. rerio</i>	851	-----	851
Pms1	<i>D. discoideum</i>	1019	QQQQ-----	1022
PMS2	<i>H. sapiens</i>	862	-----	862
Pms1	<i>S. pombe</i>	794	-----	794
Pms1	<i>A. niger</i>	844	DSLDD-----IWRGFYQEMAAHSEEEAEAE	869
Pms1	<i>S. cerevisiae</i>	873	-----	873
Pms1	<i>C. elegans</i>	805	-----	805
Pms1	<i>L. major</i>	840	-----	840
Pms1	<i>P. falciparum</i>	1295	LILSQNYDAYKYLFNHVFLLIISTKPFLLGPVLKFQ	1330
Mlh3	<i>S. cerevisiae</i>	715	-----	715
MLH3	<i>H. sapiens</i>	1424	KMAQ---AWRL--FGKAECDTRQSLQQSMPPCEPP	1453

INTERACTIONS DE LA REGION C-TERMINALE DE MLH1 NECESSAIRES A LA VOIE DE REPARATION DES MESAPPARIEMENTS DE L'ADN

La protéine Mlh1 eucaryote est un acteur central de la voie de réparation des mésappariements (MMR). Chez la levure, Mlh1 forme un hétérodimère *via* sa région C-terminale avec les endonucléases Pms1 et Mlh3. La région C-terminale de Mlh1 est également en interactions avec l'exonucléase Exo1 du MMR et deux protéines Ntg2 et Sgs1 qui sont impliquées dans d'autres voies de réparation.

Dans un premier temps, nous avons identifié et caractérisé le site d'interaction de Mlh1 avec les protéines Exo1, Ntg2 et Sgs1, qui utilisent un même motif de 5 acides aminés, (R/K)SK(Y/F)F appelé motif MIP pour Mlh1 Interacting Protein. Nous avons montré que ces 3 protéines interagissent en un même site, appelé site S2. Nous avons identifié 10 positions de Mlh1 impliquées dans le site S2 et caractérisé par microcalorimétrie, une affinité micromolaire entre des peptides contenant le motif MIP et la région C-terminale de Mlh1. Nous avons montré que les protéines EXO1 et BLM humaines qui possèdent également un motif MIP, interagissent spécifiquement avec MLH1 humain par ce motif.

Dans un second temps, nous avons résolu la structure cristallographique à 2.6Å de la région C-terminale de l'hétérodimère Mlh1*Pms1. Le site d'hétérodimérisation présente une surface d'interaction supérieure à celle observée dans les homodimères de MutL bactériens. La structure résolue confirme le rôle des 10 acides aminés de Mlh1 identifiés lors de la caractérisation du site S2. La structure du site endonucléase de Pms1 révèle la présence de deux atomes de zinc chélatés par 5 acides aminés de Pms1 et le dernier acide aminé de la protéine Mlh1, la cystéine C769. Cette première structure d'une région C-terminale d'un complexe Mlh1*Pms1 eucaryote permet d'analyser la position des nombreux mutants ponctuels de MLH1 humain associés à des cancers du côlon HNPCC.

Mots clés : Réparation des mésappariements ; MMR ; Cristallographie ; Calorimétrie ; Cancers du côlon ; Recombinaison méiotique

STRUCTURE-FUNCTION ANALYSIS OF THE INTERACTIONS MEDIATED BY MLH1 C-TERMINAL REGION AND ESSENTIAL FOR DNA MISMATCH REPAIR

Eucaryotic Mlh1 is a core component of mismatch repair pathway (MMR). In yeast organisms, Mlh1 forms heterodimer with its C-terminal region with endonucleases Pms1 and Mlh3. The C-terminal region of Mlh1 is also involved in interactions with MMR exonuclease Exo1 and two proteins, Ntg2 and Sgs1, which are involved in other DNA repair pathways.

First, we identified and charaterised the interaction site between Mlh1 and proteins Exo1, Ntg2, and Sgs1, that share the same motif of 5 amino acids, (R/K)SK(Y/F)F, named MIP box for Mlh1 Interacting Protein. We showed that these 3 proteins bind to the same site, named site S2. 10 positions of Mlh1 important for interactions on site S2 were identified and a micromolar affinity was measured by calorimetry between the C-terminal region of Mlh1 and peptides containing a MIP box. We showed that human EXO1 and BLM specifically with human MLH1 through their MIP box.

Secondly, we solved the X-ray structure of the C-terminal region of Mlh1*Pms1 heterodimer at 2.6Å. The structure shows that the surface buried upon heterodimerisation is higher in eucaryotes than in MutL homodimers. The structure confirms the overall structure of the site S2 predicted in the first part of this study. The endonuclease site of Pms1 presents in the crystal two zinc atoms that are bound by five Pms1 residues and the last residue of Mlh1 chain, cystein C769. This structure represents the first image of the C-terminal region of an eucaryote Mlh1*Pms1 heterodimer. It allows localizing the positions of human MLH1 mutants associated with colon cancers named HNPCC.

Keywords : Mismatch repair ; MMR ; Crystallography ; ITC ; Colorectal cancer ; Meiotic recombination

LABORATOIRE DE BIOLOGIE STRUCTURALE ET RADIOBIOLOGIE

iBiTec-S, SB2SM, Bâtiment 144

CEA/Saclay

91191 GIF-SUR-YVETTE Cedex

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

UNIVERSITÉ PARIS-SUD 11

UFR «FACULTÉ DE PHARMACIE DE CHATENAY-MALABRY »

5, rue Jean Baptiste Clément

92296 CHÂTENAY-MALABRY Cedex